明　細　書

音声データを選択するための装置、方法およびプログラム

技術分野

本発明は、音声データ選択装置、音声データ選択方法及びプログラムに関する。

背景技術

音声を合成する手法として、録音編集方式と呼ばれる手法がある。録音編集方式は、駅の音声案内システムや、車載用のナビゲーション装置などに用いられている。

録音編集方式は、単語と、この単語を読み上げる音声を表す音声データとを対応付けておき、音声合成する対象の文章を単語に区切ってから、これらの単語に対応付けられた音声データを取得してつなぎ合わせる、という手法である。

この録音編集方式については、例えば、特開平１０－４９１９３号公報（以降、文献１と呼ぶ）において詳細に説明されている。

しかし、音声データを単につなぎ合わせた場合、音声データ同士の境界では通常、音声のピッチ成分の周波数が不連続的に変化する、等の理由で、合成音声が不自然なものとなる。

この問題を解決する手法としては、同一の音素を互いに異なった韻律で読み上げる音声を表す複数の音声データを用意し、一方で音声合成する対象の文章に韻律予測を施して、予測結果に合致する音声データを選び出してつなぎ合わせる、という手法が考えられる。

しかし、音声データを音素毎に用意して録音編集方式により自然

な合成音声を得ようとすると、音声データを記憶する記憶装置には膨大な記憶容量が必要となり、小型軽量な装置を用いる必要がある用途には適さない。また、検索する対象のデータの量も膨大なものとなるから、高速な処理が要求される用途にも適さない。

また、韻律予測は極めて複雑な処理であるので、韻律予測を用いたこの手法を実現するには、処理能力が高いプロセッサなどを用い、あるいは長時間をかけて処理を行わせる必要がある。従ってこの手法は、構成が簡単な装置を用いた高速な処理が要求される用途には適さない。

この発明は、上記実状に鑑みてなされたものであり、簡単な構成で高速に自然な合成音声を得るための音声データ選択装置、音声データ選択方法及びプログラムを提供することを目的とする。

発明の開示

（１）上記発明目的を達成するために、本発明の音声データ選択装置は、第１の局面においては、基本的に、音声の波形を表す音声データを複数記憶する記憶手段と、文章を表す文章情報を入力し、各前記音声データのうちから、前記文章を構成する音片と読みが共通する音片の波形を表している音声データを索出する検索手段と、索出された音声データのうちから、前記文章を構成するそれぞれの音片に相当する音声データを１個ずつ、互いに隣接する音片同士の境界でのピッチの差を前記文章全体で累計した値が最小となるように選択する選択手段と、から構成される。

前記音声データ選択装置は、選択された音声データを互いに結合することにより、合成音声を表すデータを生成する音声合成手段を更に備えていてもよい。

また、本発明の音声データ選択方法は、基本的に、音声の波形を表す音声データを複数記憶し、文章を表す文章情報を入力し、各前記音声データのうちから、前記文章を構成する音片と読みが共通する音片の波形を表している音声データを索出し、および索出された音声データのうちから、前記文章を構成するそれぞれの音片に相当する音声データを１個ずつ、互いに隣接する音片同士の境界でのピッチの差を前記文章全体で累計した値が最小となるように選択する、という一連の処理ステップを含む。

また、この発明のコンピュータプログラムは、コンピュータを、音声の波形を表す音声データを複数記憶する記憶手段と、文章を表す文章情報を入力し、各前記音声データのうちから、前記文章を構成する音片と読みが共通する音片の波形を表している音声データを索出する検索手段と、索出された音声データのうちから、前記文章を構成するそれぞれの音片に相当する音声データを１個ずつ、互いに隣接する音片同士の境界でのピッチの差を前記文章全体で累計した値が最小となるように選択する選択手段と、して機能させるためのものとなっている。

（２）本発明の第２の局面においては、音声選択装置は、基本的に、音声の波形を表す音声データを複数記憶する記憶手段と、文章を表す文章情報を入力し、当該文章を構成する音片について韻律予測を行うことにより、当該音片のピッチの時間変化を予測する予測手段と、各前記音声データのうちから、前記文章を構成する音片と読みが共通する音片の波形を表していて、且つ、ピッチの時間変化が前記予測手段による予測の結果と最も高い相関を示す音声データを選択する選択手段と、から構成される。

前記選択手段は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化との間での１次回帰を行う回帰計算の結果に基づいて、当該音声データのピッチの時間変化と前記予測手段による予測の結果との相関の強さを特定するものであってもよい。

前記選択手段は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化との間の相関係数に基づいて、当該音声データのピッチの時間変化と前記予測手段による予測の結果との相関の強さを特定するものであってもよい。

また、この発明の別の音声選択装置は、音声の波形を表す音声データを複数記憶する記憶手段と、文章を表す文章情報を入力し、当該文章内の音片について韻律予測を行うことにより、当該音片の時間長、及び、当該音片のピッチの時間変化を予測する予測手段と、前記文章内の音片と読みが共通する音片の波形を表す各々の音声データについての評価値を特定し、評価値が最も高い評価を表している音声データを選択する選択手段と、より構成されており、前記評価値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果との相関を表す数値の関数、及び、当該音声データが表す音片の時間長と、当該音片と読みが共通する前記文章内の音片の時間長の予測結果との差の関数より得られるようになっている。

前記相関を表す数値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化との間での１次回帰により得られる１次関数の勾配からなって

いてもよい。

また、前記相関を表す数値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化との間での１次回帰により得られる１次関数の切片からなっていてもよい。

前記相関を表す数値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果との間の相関係数からなっていてもよい。

前記相関を表す数値は、音声データが表す音片のピッチの時間変化を表すデータを種々のビット数循環シフトしたものが表す関数と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果を表す関数との相関係数の最大値からなっていてもよい。

前記記憶手段は、音声データの読みを表す表音データを、当該音声データに対応付けて記憶していてもよく、また前記選択手段は、前記文章内の音片の読みに合致する読みを表す表音データが対応付けられている音声データを、当該音片と読みが共通する音片の波形を表す音声データとして扱うものであってもよい。

前記音声選択装置は、選択された音声データを互いに結合することにより、合成音声を表すデータを生成する音声合成手段を更に備えていてもよい。

前記音声選択装置は、前記文章内の音片のうち、前記選択手段が音声データを選択できなかった音片について、前記記憶手段が記憶する音声データを用いることなく、当該音片の波形を表す音声データを合成する欠落部分合成手段を備えていてもよく、前記音声合成手段は、前記選択手段が選択した音声データ及び前記欠落部分合成

手段が合成した音声データを互いに結合することにより、合成音声を表すデータを生成するものであってもよい。

また、この発明の音声選択方法は、音声の波形を表す音声データを複数記憶し、文章を表す文章情報を入力し、当該文章を構成する音片について韻律予測を行うことにより、当該音片のピッチの時間変化を予測し、および各前記音声データのうちから、前記文章を構成する音片と読みが共通する音片の波形を表していて、且つ、ピッチの時間変化が前記予測手段による予測の結果と最も高い相関を示す音声データを選択する、という一連の処理ステップを含む。

また、この発明の別の音声選択方法は、音声の波形を表す音声データを複数記憶し、文章を表す文章情報を入力し、当該文章内の音片について韻律予測を行うことにより、当該音片の時間長、及び、当該音片のピッチの時間変化を予測し、前記文章内の音片と読みが共通する音片の波形を表す各々の音声データについての評価値を特定し、評価値が最も高い評価を表している音声データを選択するようになっており、前記評価値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果との相関を表す数値の関数、及び、当該音声データが表す音片の時間長と、当該音片と読みが共通する前記文章内の音片の時間長の予測結果との差の関数より得られるものである。

また、この発明のコンピュータプログラムは、コンピュータを、音声の波形を表す音声データを複数記憶する記憶手段と、文章を表す文章情報を入力し、当該文章を構成する音片について韻律予測を行うことにより、当該音片のピッチの時間変化を予測する予測手段と、各前記音声データのうちから、前記文章を構成する音片と読み

が共通する音片の波形を表していて、且つ、ピッチの時間変化が前記予測手段による予測の結果と最も高い相関を示す音声データを選択する選択手段と、して機能させるためのものとなっている。

さらに、この発明の別のコンピュータプログラムは、コンピュータを、音声の波形を表す音声データを複数記憶する記憶手段と、文章を表す文章情報を入力し、当該文章内の音片について韻律予測を行うことにより、当該音片の時間長、及び、当該音片のピッチの時間変化を予測する予測手段と、前記文章内の音片と読みが共通する音片の波形を表す各々の音声データについての評価値を特定し、評価値が最も高い評価を表している音声データを選択する選択手段として機能させるためのプログラムであって、前記評価値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果との相関を表す数値の関数、及び、当該音声データが表す音片の時間長と、当該音片と読みが共通する前記文章内の音片の時間長の予測結果との差の関数より得られるものである、となっている。

（３）本発明の第３の局面においては、音声データ選択装置は、基本的に、音声の波形を表す音声データを複数記憶する記憶手段と、文章を表す文章情報を入力する文章情報入力手段と、前記文章情報が表す文章内の音片と読みが共通する部分を有する音声データを索出する検索部と、前記索出されたそれぞれの音声データを文章情報が表す文章に従って接続した際に互いに隣接する音声データ同士の関係に基づいた所定の評価基準に従って評価値を求め、出力する音声データの組み合わせを当該評価値に基づいて選択する選択手段と、から構成される。

前記評価基準は、音声データが表す音声と韻律予測結果との相関及び互いに隣接する音声データ同士の関係を示す評価値を定める基準であって、前記評価値は、前記音声データが表す音声の特徴を示すパラメータ、前記音声データが表す音声を互いに結合して得られる音声の特徴を示すパラメータ、及び、発話時間長に関する特徴を示すパラメータのうち、少なくともいずれかを含む評価式に基づいて得られるものであるものであってもよい。

あるいは、前記評価基準は、音声データが表す音声と韻律予測結果との相関及び互いに隣接する音声データ同士の関係を示す評価値を定める基準であって、前記評価値は、前記音声データが表す音声を互いに結合して得られる音声の特徴を示すパラメータを含み、また、前記音声データが表す音声の特徴を示すパラメータと発話時間長に関する特徴を示すパラメータのうち、少なくともいずれかを含む評価式に基づいて得られるものであってもよい。

前記音声データが表す音声を互いに結合して得られる音声の特徴を示すパラメータは、前記文章情報が表す文章内の音片と読みが共通する部分を有する音声の波形を表す音声データのうちから、前記文章を構成するそれぞれの音片に相当する音声データを１個ずつ選択した場合における、互いに隣接する音声データ同士の境界でのピッチの差に基づいて得られるものであってもよい。

前記音片データ選択装置は、文章を表す文章情報を入力し、当該文章内の音片について韻律予測を行うことにより、当該音片の時間長、及び、当該音片のピッチの時間変化を予測する予測手段を備えていてもよく、前記評価基準は、音声データが表す音声と前記韻律予測手段の韻律予測結果との相関ないし差異を示す評価値を定める

基準であって、前記評価値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果との相関を表す数値の関数、及び／又は、当該音声データが表す音片の時間長と、当該音片と読みが共通する前記文章内の音片の時間長の予測結果との差の関数に基づいて得られるものであってもよい。

前記相関を表す数値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化との間での１次回帰により得られる１次関数の勾配及び／又は切片からなっていてもよい。

前記相関を表す数値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果との間の相関係数からなっていてもよい。

あるいは、前記相関を表す数値は、音声データが表す音片のピッチの時間変化を表すデータを種々のビット数循環シフトしたものが表す関数と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果を表す関数との相関係数の最大値からなっていてもよい。

前記記憶手段は、音声データの読みを表す表音データを、当該音声データに対応付けて記憶していてもよく、前記選択手段は、前記文章内の音片の読みに合致する読みを表す表音データが対応付けられている音声データを、当該音片と読みが共通する音片の波形を表す音声データとして扱うものであってもよい。

前記音片データ選択装置は、選択された音声データを互いに結合することにより、合成音声を表すデータを生成する音声合成手段を

更に備えていてもよい。

前記音片データ選択装置は、前記文章内の音片のうち、前記選択手段が音声データを選択できなかった音片について、前記記憶手段が記憶する音声データを用いることなく、当該音片の波形を表す音声データを合成する欠落部分合成手段を備えていてもよく、前記音声合成手段は、前記選択手段が選択した音声データ及び前記欠落部分合成手段が合成した音声データを互いに結合することにより、合成音声を表すデータを生成するものであってもよい。

また、この発明の音声データ選択方法は、音声の波形を表す音声データを複数記憶し、文章を表す文章情報を入力し、前記文章情報が表す文章内の音片と読みが共通する部分を有する音声データを索出し、前記索出されたそれぞれの音声データを文章情報が表す文章に従って接続した際に互いに隣接する音声データ同士の関係に基づいた所定の評価基準に従って評価値を求め、出力する音声データの組み合わせを当該評価値に基づいて選択する、一連の処理ステップを含む。

また、この発明のコンピュータプログラムは、コンピュータを、音声の波形を表す音声データを複数記憶する記憶手段と、文章を表す文章情報を入力する文章情報入力手段と、前記文章情報が表す文章内の音片と読みが共通する部分を有する音声データを索出する検索部と、前記索出されたそれぞれの音声データを文章情報が表す文章に従って接続した際に互いに隣接する音声データ同士の関係に基づいた所定の評価基準に従って評価値を求め、出力する音声データの組み合わせを当該評価値に基づいて選択する選択手段と、して機能させるためのものとなっている。

図面の簡単な説明

第１図は、この発明の各実施の形態に係る音声合成システムの構成を示すブロック図である。

第２図は、この発明の第１の実施の形態における音片データベースのデータ構造を模式的に示す図である。

第３図の（ａ）は、音片についてのピッチ成分の周波数の予測結果と、この音片と読みが合致する音片の波形を表す音片データのピッチ成分の周波数の時間変化とを１次回帰させる処理を説明するためのグラフであり、同図（ｂ）は、相関係数を求めるために用いる予測結果データ及びピッチ成分データの値の一例を示すグラフである。

第４図は、この発明の第２の実施の形態における音片データベースのデータ構造を模式的に示す図である。

第５図の（ａ）は、定型メッセージの読みを示す図であり、同図（ｂ）は、音片編集部に供給された音片データのリストであり、同図（ｃ）は、先行する音片の末尾におけるピッチ成分の周波数と後続の音片の先頭におけるピッチ成分の周波数との差の絶対値を示す図であり、同図（ｄ）は、音片編集部がどの音片データを選択するかを示す図である。

第６図は、この発明の各実施の形態に係る音声合成システムの機能を行うパーソナルコンピュータがフリーテキストデータを取得した場合の処理を示すフローチャートである。

第７図は、この発明の各実施の形態に係る音声合成システムの機能を行うパーソナルコンピュータが配信文字列データを取得した場合の処理を示すフローチャートである。

第８図は、この発明の第１の実施の形態に係る音声合成システムの機能を行うパーソナルコンピュータが定型メッセージデータ及び発声スピードデータを取得した場合の処理を示すフローチャートである。

第９図は、この発明の第２の実施の形態に係る音声合成システムの機能を行うパーソナルコンピュータが定型メッセージデータ及び発声スピードデータを取得した場合の処理を示すフローチャートである。

第１０図は、この発明の第３の実施の形態に係る音声合成システムの機能を行うパーソナルコンピュータが定型メッセージデータ及び発声スピードデータを取得した場合の処理を示すフローチャートである。

発明を実施するための最良の形態

以下、この発明の実施の形態を、音声合成システムを例とし、図面を参照して説明する。

（第１の実施の形態）

第１図は、この発明の第１の実施の形態に係る音声合成システムの構成を示す図である。図示するように、この音声合成システムは、本体ユニットＭと、音片登録ユニットＲとにより構成されている。

本体ユニットＭは、言語処理部１と、一般単語辞書２と、ユーザ単語辞書３と、音響処理部４と、検索部５と、伸長部６と、波形データベース７と、音片編集部８と、検索部９と、音片データベース１０と、話速変換部１１とにより構成されている。

言語処理部１、音響処理部４、検索部５、伸長部６、音片編集部８、検索部９及び話速変換部１１は、いずれも、ＣＰＵ（Central

Processing Unit）やＤＳＰ（Digital Signal Processor）等のプロセッサや、このプロセッサが実行するためのプログラムを記憶するメモリなどより構成されており、それぞれ後述する処理を行う。

なお、言語処理部１、音響処理部４、検索部５、伸長部６、音片編集部８、検索部９及び話速変換部１１の一部又は全部の機能を単一のプロセッサが行うようにしてもよい。

一般単語辞書２は、ＰＲＯＭ（Programmable Read Only Memory）やハードディスク装置等の不揮発性メモリより構成されている。一般単語辞書２には、表意文字（例えば、漢字など）を含む単語等と、この単語等の読みを表す表音文字（例えば、カナや発音記号など）とが、この音声合成システムの製造者等によって、あらかじめ互いに対応付けて記憶されている。

ユーザ単語辞書３は、ＥＥＰＲＯＭ（Electrically Erasable/Programmable Read Only Memory）やハードディスク装置等のデータ書き換え可能な不揮発性メモリと、この不揮発性メモリへのデータの書き込みを制御する制御回路とにより構成されている。なお、プロセッサがこの制御回路の機能を行ってもよく、言語処理部１、音響処理部４、検索部５、伸長部６、音片編集部８、検索部９及び話速変換部１１の一部又は全部の機能を行うプロセッサがユーザ単語辞書３の制御回路の機能を行うようにしてもよい。

ユーザ単語辞書３は、表意文字を含む単語等と、この単語等の読みを表す表音文字とを、ユーザの操作に従って外部より取得し、互いに対応付けて記憶する。ユーザ単語辞書３には、一般単語辞書２に記憶されていない単語等とその読みを表す表音文字とが格納され

ていれば十分である。

波形データベース７は、ＰＲＯＭやハードディスク装置等の不揮発性メモリより構成されている。波形データベース７には、表音文字と、この表音文字が表す単位音声の波形を表す波形データをエントロピー符号化して得られる圧縮波形データとが、この音声合成システムの製造者等によって、あらかじめ互いに対応付けて記憶されている。単位音声は、規則合成方式の手法で用いられる程度の短い音声であり、具体的には、音素や、ＶＣＶ（Vowel-Consonant-Vowel）音節などの単位で区切られる音声である。なお、エントロピー符号化される前の波形データは、例えば、ＰＣＭ（Pulse Code Modulation）化されたデジタル形式のデータからなっていればよい。

音片データベース１０は、ＰＲＯＭやハードディスク装置等の不揮発性メモリより構成されている。

音片データベース１０には、例えば、第２図に示すデータ構造を有するデータが記憶されている。すなわち、図示するように、音片データベース１０に格納されているデータは、ヘッダ部ＨＤＲ、インデックス部ＩＤＸ、ディレクトリ部ＤＩＲ及びデータ部ＤＡＴの４種に分かれている。

なお、音片データベース１０へのデータの格納は、例えば、この音声合成システムの製造者によりあらかじめ行われ、及び／又は、音片登録ユニットＲが後述する動作を行うことにより行われる。

ヘッダ部ＨＤＲには、音片データベース１０を識別するデータや、インデックス部ＩＤＸ、ディレクトリ部ＤＩＲ及びデータ部ＤＡＴのデータ量、データの形式、著作権等の帰属などを示すデータが格

納される。

データ部ＤＡＴには、音片の波形を表す音片データをエントロピー符号化して得られる圧縮音片データが格納されている。

なお、音片とは、音声のうち音素１個以上を含む連続した１区間をいい、通常は単語１個分又は複数個分の区間からなる。

また、エントロピー符号化される前の音片データは、上述の圧縮波形データの生成のためエントロピー符号化される前の波形データと同じ形式のデータ（例えば、ＰＣＭされたデジタル形式のデータ）からなっていればよい。

ディレクトリ部ＤＩＲには、個々の圧縮音声データについて、

（Ａ）　この圧縮音片データが表す音片の読みを示す表音文字を表すデータ（音片読みデータ）、

（Ｂ）　この圧縮音片データが格納されている記憶位置の先頭のアドレスを表すデータ、

（Ｃ）　この圧縮音片データのデータ長を表すデータ、

（Ｄ）　この圧縮音片データが表す音片の発声スピード（再生した場合の時間長）を表すデータ（スピード初期値データ）、

（Ｅ）　この音片のピッチ成分の周波数の時間変化を表すデータ(ピッチ成分データ）、

が、互いに対応付けられた形で格納されている。（なお、音片データベース１０の記憶領域にはアドレスが付されているものとする。）

なお、第２図は、データ部ＤＡＴに含まれるデータとして、読みが「サイタマ」である音片の波形を表す、データ量１４１０ｈバイトの圧縮音片データが、アドレス００１Ａ３６Ａ６ｈを先頭とする

論理的位置に格納されている場合を例示している。（なお、本明細書及び図面において、末尾に“ｈ”を付した数字は１６進数を表す。）

また、ピッチ成分データは、例えば、図示するように、音片のピッチ成分の周波数をサンプリングして得られたサンプルＹ（ｉ）（サンプルの総数をｎとして、ｉはｎ以下の正の整数）を表すデータであるものとする。

なお、上述の（Ａ）～（Ｅ）のデータの集合のうち少なくとも（Ａ）のデータ（すなわち音片読みデータ）は、音片読みデータが表す表音文字に基づいて決められた順位に従ってソートされた状態で（例えば、表音文字がカナであれば、五十音順に従って、アドレス降順に並んだ状態で）、音片データベース１０の記憶領域に格納されている。

インデックス部ＩＤＸには、ディレクトリ部ＤＩＲのデータのおおよその論理的位置を音片読みデータに基づいて特定するためのデータが格納されている。具体的には、例えば、音片読みデータがカナを表すものであるとして、カナ文字と、先頭１字がこのカナ文字であるような音片読みデータがどのような範囲のアドレスにあるかを示すデータとが、互いに対応付けて格納されている。

なお、一般単語辞書２、ユーザ単語辞書３、波形データベース７及び音片データベース１０の一部又は全部の機能を単一の不揮発性メモリが行うようにしてもよい。

音片データベース１０へのデータの格納は、第１図に示す音片登録ユニットＲにより行われる。音片登録ユニットＲは、図示するように、収録音片データセット記憶部１２と、音片データベース作成部１３と、圧縮部１４とにより構成されている。なお、音片登録ユ

ニットＲは音片データベース１０とは着脱可能に接続されていてもよく、この場合は、音片データベース１０に新たにデータを書き込むときを除いては、音片登録ユニットＲを本体ユニットＭから切り離した状態で本体ユニットＭに後述の動作を行わせてよい。

収録音片データセット記憶部１２は、ハードディスク装置等のデータ書き換え可能な不揮発性メモリより構成されている。

収録音片データセット記憶部１２には、音片の読みを表す表音文字と、この音片を人が実際に発声したものを集音して得た波形を表す音片データとが、この音声合成システムの製造者等によって、あらかじめ互いに対応付けて記憶されている。なお、この音片データは、例えば、ＰＣＭ化されたデジタル形式のデータからなっていればよい。

音片データベース作成部１３及び圧縮部１４は、ＣＰＵ等のプロセッサや、このプロセッサが実行するためのプログラムを記憶するメモリなどより構成されており、このプログラムに従って後述する処理を行う。

なお、音片データベース作成部１３及び圧縮部１４の一部又は全部の機能を単一のプロセッサが行うようにしてもよく、また、言語処理部１、音響処理部４、検索部５、伸長部６、音片編集部８、検索部９及び話速変換部１１の一部又は全部の機能を行うプロセッサが音片データベース作成部１３や圧縮部１４の機能を更に行ってもよい。また、音片データベース作成部１３や圧縮部１４の機能を行うプロセッサが、収録音片データセット記憶部１２の制御回路の機能を兼ねてもよい。

音片データベース作成部１３は、収録音片データセット記憶部１

２より、互いに対応付けられている表音文字及び音片データを読み出し、この音片データが表す音声のピッチ成分の周波数の時間変化と、発声スピードとを特定する。

発声スピードの特定は、例えば、この音片データのサンプル数を数えることにより特定すればよい。

一方、ピッチ成分の周波数の時間変化は、例えば、この音片データにケプストラム解析を施すことにより特定すればよい。具体的には、例えば、音片データが表す波形を時間軸上で多数の小部分へと区切り、得られたそれぞれの小部分の強度を、元の値の対数（対数の底は任意）に実質的に等しい値へと変換し、値が変換されたこの小部分のスペクトル（すなわち、ケプストラム）を、高速フーリエ変換の手法（あるいは、離散的変数をフーリエ変換した結果を表すデータを生成する他の任意の手法）により求める。そして、このケプストラムの極大値を与える周波数のうちの最小値を、この小部分におけるピッチ成分の周波数として特定する。

なお、ピッチ成分の周波数の時間変化は、例えば、特開２００３－１０８１７２号公報に開示された手法に従って音片データをピッチ波形データへと変換してから、このピッチ波形データに基づいて特定するようにすると良好な結果が期待できる。具体的には、音片データをフィルタリングしてピッチ信号を抽出し、抽出されたピッチ信号に基づいて、音片データが表す波形を単位ピッチ長の区間へと区切り、各区間について、ピッチ信号との相関関係に基づいて位相のずれを特定して各区間の位相を揃えることにより、音片データをピッチ波形信号へと変換すればよい。そして、得られたピッチ波形信号を音片データとして扱い、ケプストラム解析を行う等するこ

とにより、ピッチ成分の周波数の時間変化を特定すればよい。

一方、音片データベース作成部１３は、収録音片データセット記憶部１２より読み出した音片データを圧縮部１４に供給する。

圧縮部１４は、音片データベース作成部１３より供給された音片データをエントロピー符号化して圧縮音片データを作成し、音片データベース作成部１３に返送する。

音片データの発声スピード及びピッチ成分の周波数の時間変化を特定し、この音片データがエントロピー符号化され圧縮音片データとなって圧縮部１４より返送されると、音片データベース作成部１３は、この圧縮音片データを、データ部ＤＡＴを構成するデータとして、音片データベース１０の記憶領域に書き込む。

また、音片データベース作成部１３は、書き込んだ圧縮音片データが表す音片の読みを示すものとして収録音片データセット記憶部１２より読み出した表音文字を、音片読みデータとして音片データベース１０の記憶領域に書き込む。

また、書き込んだ圧縮音片データの、音片データベース１０の記憶領域内での先頭のアドレスを特定し、このアドレスを上述の（Ｂ）のデータとして音片データベース１０の記憶領域に書き込む。

また、この圧縮音片データのデータ長を特定し、特定したデータ長を、（Ｃ）のデータとして音片データベース１０の記憶領域に書き込む。

また、この圧縮音片データが表す音片の発声スピード及びピッチ成分の周波数の時間変化を特定した結果を示すデータを生成し、スピード初期値データ及びピッチ成分データとして音片データベース１０の記憶領域に書き込む。

次に、この音声合成システムの動作を説明する。

まず、言語処理部１が、この音声合成システムに音声を合成させる対象としてユーザが用意した、表意文字を含む文章（フリーテキスト）を記述したフリーテキストデータを外部から取得したとして説明する。

なお、言語処理部１がフリーテキストデータを取得する手法は任意であり、例えば、図示しないインターフェース回路を介して外部の装置やネットワークから取得してもよいし、図示しない記録媒体ドライブ装置にセットされた記録媒体（例えば、フロッピー（登録商標）ディスクやＣＤ－ＲＯＭなど）から、この記録媒体ドライブ装置を介して読み取ってもよい。また、言語処理部１の機能を行っているプロセッサが、自ら実行している他の処理で用いたテキストデータを、フリーテキストデータとして、言語処理部１の処理へと引き渡すようにしてもよい。

フリーテキストデータを取得すると、言語処理部１は、このフリーテキストに含まれるそれぞれの表意文字について、その読みを表す表音文字を、一般単語辞書２やユーザ単語辞書３を検索することにより特定する。そして、この表意文字を、特定した表音文字へと置換する。そして、言語処理部１は、フリーテキスト内の表意文字がすべて表音文字へと置換した結果得られる表音文字列を、音響処理部４へと供給する。

音響処理部４は、言語処理部１より表音文字列を供給されると、この表音文字列に含まれるそれぞれの表音文字について、当該表音文字が表す単位音声の波形を検索するよう、検索部５に指示する。

検索部５は、この指示に応答して波形データベース７を検索し、

表音文字列に含まれるそれぞれの表音文字が表す単位音声の波形を表す圧縮波形データを索出する。そして、索出された圧縮波形データを伸長部６へと供給する。

伸長部６は、検索部５より供給された圧縮波形データを、圧縮される前の波形データへと復元し、検索部５へと返送する。検索部５は、伸長部６より返送された波形データを、検索結果として音響処理部４へと供給する。

音響処理部４は、検索部５より供給された波形データを、言語処理部１より供給された表音文字列内での各表音文字の並びに従った順序で、音片編集部８へと供給する。

音片編集部８は、音響処理部４より波形データを供給されると、この波形データを、供給された順序で互いに結合し、合成音声を表すデータ（合成音声データ）として出力する。フリーテキストデータに基づいて合成されたこの合成音声は、規則合成方式の手法により合成された音声に相当する。

なお、音片編集部８が合成音声データを出力する手法は任意であり、例えば、図示しないＤ／Ａ（Digital-to-Analog）変換器やスピーカを介して、この合成音声データが表す合成音声を再生するようにしてもよい。また、図示しないインターフェース回路を介して外部の装置やネットワークに送出してもよいし、図示しない記録媒体ドライブ装置にセットされた記録媒体へ、この記録媒体ドライブ装置を介して書き込んでもよい。また、音片編集部８の機能を行っているプロセッサが、自ら実行している他の処理へと、合成音声データを引き渡すようにしてもよい。

次に、音響処理部４が、外部より配信された、表音文字列を表す

データ（配信文字列データ）を取得したとする。（なお、音響処理部４が配信文字列データを取得する手法も任意であり、例えば、言語処理部１がフリーテキストデータを取得する手法と同様の手法で配信文字列データを取得すればよい。）

この場合、音響処理部４は、配信文字列データが表す表音文字列を、言語処理部１より供給された表音文字列と同様に扱う。この結果、配信文字列データが表す表音文字列に含まれる表音文字に対応する圧縮波形データが検索部５により索出され、圧縮される前の波形データが伸長部６により復元される。復元された各波形データは音響処理部４を介して音片編集部８へと供給され、音片編集部８が、この波形データを、配信文字列データが表す表音文字列内での各表音文字の並びに従った順序で互いに結合し、合成音声データとして出力する。配信文字列データに基づいて合成されたこの合成音声データも、規則合成方式の手法により合成された音声を表す。

次に、音片編集部８が、定型メッセージデータ及び発声スピードデータを取得したとする。

なお、定型メッセージデータは、定型メッセージを表音文字列として表すデータであり、発声スピードデータは、定型メッセージデータが表す定型メッセージの発声スピードの指定値（この定型メッセージを発声する時間長の指定値）を示すデータである。

また、音片編集部８が定型メッセージデータや発声スピードデータを取得する手法は任意であり、例えば、言語処理部１がフリーテキストデータを取得する手法と同様の手法で定型メッセージデータや発声スピードデータを取得すればよい。

定型メッセージデータ及び発声スピードデータが音片編集部８に

供給されると、音片編集部８は、定型メッセージに含まれる音片の読みを表す表音文字に合致する表音文字が対応付けられている圧縮音片データをすべて索出するよう、検索部９に指示する。

検索部９は、音片編集部８の指示に応答して音片データベース１０を検索し、該当する圧縮音片データと、該当する圧縮音片データに対応付けられている上述の音片読みデータ、スピード初期値データ及びピッチ成分データとを索出し、索出された圧縮波形データを伸長部６へと供給する。１個の音片につき複数の圧縮音片データが該当する場合も、該当する圧縮音片データすべてが、音声合成に用いられるデータの候補として索出される。一方、圧縮音片データを索出できなかった音片があった場合、検索部９は、該当する音片を識別するデータ（以下、欠落部分識別データと呼ぶ）を生成する。

伸長部６は、検索部９より供給された圧縮音片データを、圧縮される前の音片データへと復元し、検索部９へと返送する。検索部９は、伸長部６より返送された音片データと、索出された音片読みデータ、スピード初期値データ及びピッチ成分データとを、検索結果として話速変換部１１へと供給する。また、欠落部分識別データを生成した場合は、この欠落部分識別データも話速変換部１１へと供給する。

一方、音片編集部８は、話速変換部１１に対し、話速変換部１１に供給された音片データを変換して、当該音片データが表す音片の時間長を、発声スピードデータが示すスピードに合致するようにすることを指示する。

話速変換部１１は、音片編集部８の指示に応答し、検索部９より供給された音片データを指示に合致するように変換して、音片編集

部８に供給する。具体的には、例えば、検索部９より供給された音片データの元の時間長を、索出されたスピード初期値データに基づいて特定した上、この音片データをリサンプリングして、この音片データのサンプル数を、音片編集部８の指示したスピードに合致する時間長にすればよい。

また、話速変換部１１は、検索部９より供給された音片読みデータ、スピード初期値データ及びピッチ成分データも音片編集部８に供給し、欠落部分識別データを検索部９より供給された場合は、更にこの欠落部分識別データも音片編集部８に供給する。

なお、発声スピードデータが音片編集部８に供給されていない場合、音片編集部８は、話速変換部１１に対し、話速変換部１１に供給された音片データを変換せずに音片編集部８に供給するよう指示すればよく、話速変換部１１は、この指示に応答し、検索部９より供給された音片データをそのまま音片編集部８に供給すればよい。

音片編集部８は、話速変換部１１より音片データ、音片読みデータ、スピード初期値データ及びピッチ成分データを供給されると、供給された音片データのうちから、定型メッセージを構成する音片の波形に最もよく近似できる波形を表す音片データを、音片１個につき１個ずつ選択する。

具体的には、まず、音片編集部８は、定型メッセージデータが表す定型メッセージに、例えば「藤崎モデル」や「ＴｏＢＩ（Tone and Break Indices）」等の韻律予測の手法に基づいた解析を加えることにより、この定型メッセージ内の各音片のピッチ成分の周波数の時間変化を予測する。そして、音片毎に、ピッチ成分の周波数の時間変化の予測結果をサンプリングしたものを表すデジタル形式のデ

ータ（以下、予測結果データと呼ぶ）を生成する。

次に、音片編集部８は、定型メッセージ内のそれぞれの音片について、この音片のピッチ成分の周波数の時間変化の予測結果を表す予測結果データと、この音片と読みが合致する音片の波形を表す音片データのピッチ成分の周波数の時間変化を表すピッチ成分データとの相関を求める。

より具体的には、音片編集部８は、話速変換部１１より供給された各々のピッチ成分データについて、例えば、数式１の右辺に示す値α及び数式２の右辺に示す値βを求める。

$$\alpha = \frac{\sum_{i=1}^{n} [\{X(i) - mx\} \cdot \{Y(i) - my\}]}{\sum_{i=1}^{n} \{X(i) - mx\}^2}$$

ただし、$mx = \sum_{i=1}^{n} \frac{X(i)}{n}$, $my = \sum_{i=1}^{n} \frac{Y(i)}{n}$

$$\beta = my - (\alpha \cdot mx)$$

第３図（ａ）に示すように、ある音片についての予測結果データ（サンプルの総数はｎ個とする）のｉ番目のサンプルの値Ｘ（ｉ）（ｉは整数）の１次関数として、この音片と読みが合致する音片の波形を表す音片データについてのピッチ成分データ（サンプルの総数はｎ個とする）のｉ番目のサンプルＹ（ｉ）の値を１次回帰させ

た場合、この１次関数の勾配はα、切片はβとなる。（勾配αの単位は例えば［ヘルツ／秒］であればよく、切片βの単位は例えば［ヘルツ］であればよい。）

なお、同一の読みの音片について、予測結果データとピッチ成分データとでサンプルの総数が互いに異なる場合は、両者のうち一方（または両方）を、１次補間やラグランジェ補間あるいはその他任意の手法により補間した上でリサンプリングし、両者のサンプルの総数を揃えてから相関を求めるようにすればよい。

一方、音片編集部８は、話速変換部１１より供給されたスピード初期値データと、音片編集部８に供給された定型メッセージデータ及び発声スピードデータとを用いて、数式３の右辺の値ｄｔを求める。この値ｄｔは、音片データが表す音片の発声スピードと、この音片と読みが合致する定型メッセージ内の音片の発声スピードとの時間差を表す係数である。

$$dt = |(Xt - Yt)/Yt|$$

（ただし、Ｙｔは音片データが表す音片の発声スピード、Ｘｔはこの音片と読みが合致する定型メッセージ内の音片の発声スピード）

そして、音片編集部８は、１次回帰により得られた上述のα及びβの値と、上述の係数ｄｔとに基づいて、定型メッセージ内の音片の読みと一致する音片を表す音片データのうち、数式４の右辺の値（評価値）ｃｏｓｔ１が最大となるものを選択する。

$$cost1 = 1/(W_1|1-\alpha| + W_2|\beta| + dt)$$

（ただし、W_1及びW_2は所定の正の係数）

音片のピッチ成分の周波数の時間変化の予測結果と、この音片と読みが合致する音片の波形を表す音片データのピッチ成分の周波数

の時間変化とが互いに近いほど、勾配αの値は1に近くなり、従って、値｜1－α｜は0に近くなる。そして、評価値ｃｏｓｔ1は、音片のピッチの予測結果と音片データのピッチとの相関が高いほど大きな値となるようにするため、値｜1－α｜の1次関数の逆数の形をとっているので、評価値ｃｏｓｔ1は、値｜1－α｜が0に近くなるほど大きな値となる。

一方、音声の抑揚は、音片のピッチ成分の周波数の時間変化により特徴付けられる。従って、勾配αの値は、音声の抑揚の差異を敏感に反映する性質を有する。

このため、合成されるべき音声について抑揚の正確さが重視される場合（例えば、電子メール等のテキストを読み上げる音声を合成する場合等）は、上述の係数W_1の値をなるべく大きくすることが望ましい。

これに対し、音片のピッチ成分の基本周波数（ベースピッチ周波数）の予測結果と、この音片と読みが合致する音片の波形を表す音片データのベースピッチ周波数とが互いに近いほど、切片βの値は0に近くなる。従って、切片βの値は、音声のベースピッチ周波数の差異を敏感に反映する性質を有する。一方、評価値ｃｏｓｔ1は、値｜β｜の1次関数の逆数とみることもできる形をとっているので、評価値ｃｏｓｔ1は、値｜β｜が0に近くなるほど大きな値となる。

一方、音声のベースピッチ周波数は、音声の話者の声質を支配する要因であり、話者の性別による差異も顕著である。

このため、合成されるべき音声についてベースピッチ周波数の正確さが重視される場合（例えば、合成音声の話者の性別や声質を明確にする必要がある場合など）は、上述の係数W_2の値をなるべく

大きくすることが望ましい。

動作の説明に戻ると、音片編集部８は、定型メッセージ内の音片の波形に近い波形を表す音片データを選択する一方で、話速変換部１１より欠落部分識別データも供給されている場合には、欠落部分識別データが示す音片の読みを表す表音文字列を定型メッセージデータより抽出して音響処理部４に供給し、この音片の波形を合成するよう指示する。

指示を受けた音響処理部４は、音片編集部８より供給された表音文字列を、配信文字列データが表す表音文字列と同様に扱う。この結果、この表音文字列に含まれる表音文字が示す音声の波形を表す圧縮波形データが検索部５により索出され、この圧縮波形データが伸長部６により元の波形データへと復元され、検索部５を介して音響処理部４へと供給される。音響処理部４は、この波形データを音片編集部８へと供給する。

音片編集部８は、音響処理部４より波形データを返送されると、この波形データと、話速変換部１１より供給された音片データのうち音片編集部８が特定したものとを、定型メッセージデータが示す定型メッセージ内での各音片の並びに従った順序で互いに結合し、合成音声を表すデータとして出力する。

なお、話速変換部１１より供給されたデータに欠落部分識別データが含まれていない場合は、音響処理部４に波形の合成を指示することなく直ちに、音片編集部８が特定した音片データを、定型メッセージデータが示す定型メッセージ内での各音片の並びに従った順序で互いに結合し、合成音声を表すデータとして出力すればよい。

以上説明した、この音声合成システムでは、音素より大きな単位

であり得る音片の波形を表す音片データが、韻律の予測結果に基づいて、録音編集方式により自然につなぎ合わせられ、定型メッセージを読み上げる音声が合成される。音片データベース１０の記憶容量は、音素毎に波形を記憶する場合に比べて小さくでき、また、高速に検索できる。このため、この音声合成システムは小型軽量に構成することができ、また高速な処理にも追随できる。

また、音片の波形の予測結果と音片データとの相関を複数の評価基準（例えば、１次回帰させた場合の勾配や切片による評価と、音片の時間差による評価、など）で評価した場合は、これらの評価の結果に食い違いが生じる場合が多々あり得る。しかし、この音声合成システムでは、複数の評価基準で評価した結果が１個の評価値に基づいて総合され、適正な評価が行われる。

なお、この音声合成システムの構成は上述のものに限られない。

例えば、波形データや音片データはＰＣＭ形式のデータである必要はなく、データ形式は任意である。

また、波形データベース７や音片データベース１０は波形データや音片データを必ずしもデータ圧縮された状態で記憶している必要はない。波形データベース７や音片データベース１０が波形データや音片データをデータ圧縮されていない状態で記憶している場合、本体ユニットＭは伸長部６を備えている必要はない。

また、音片データベース作成部１３は、図示しない記録媒体ドライブ装置にセットされた記録媒体から、この記録媒体ドライブ装置を介して、音片データベース１０に追加する新たな圧縮音片データの材料となる音片データや表音文字列を読み取ってもよい。

また、音片登録ユニットＲは、必ずしも収録音片データセット記

憶部１２を備えている必要はない。

また、音片編集部８は、特定の音片の韻律を表す韻律登録データをあらかじめ記憶し、定型メッセージにこの特定の音片が含まれている場合は、この韻律登録データが表す韻律を、韻律予測の結果として扱うようにしてもよい。

また、音片編集部８は、過去の韻律予測の結果を韻律登録データとして新たに記憶するようにしてもよい。

また、音片編集部８は、上述のα及びβの値を求める代わりに、話速変換部１１より供給された各々のピッチ成分データについて、例えば、数式５の右辺に示す値Ｒｘｙ（ｊ）を、ｊの値を０以上ｎ未満の各整数として、合計ｎ個求め、得られたＲｘｙ（０）からＲｘｙ（ｎ－１）までのｎ個の相関係数のうちの最大値を特定するようにしてもよい。

$$Rxy(j)=\frac{\sum_{i=1}^{n}[\{X(i)-mx\}\cdot\{Yj(i)-my\}]}{\sqrt{\sum_{i=1}^{n}\{X(i)-mx\}^{2}}\sqrt{\sum_{i=1}^{n}\{Y(i)-my\}^{2}}}$$

Ｒｘｙ（ｊ）は、ある音片についての予測結果データ（サンプル総数ｎ個。なお、数式５におけるＸ（ｉ）は数式１におけるものと同一である）と、この音片と読みが合致する音片の波形を表す音片データについてのピッチ成分データ（サンプルの総数ｎ個）を一定の方向へｊ個循環シフトして得られたサンプルの列（なお、数式５においてＹｊ（ｉ）は、このサンプルの列のｉ番目のサンプルの値である）との相関係数の値である。

なお、第３図（ｂ）は、Ｒｘｙ（０）及びＲｘｙ（ｊ）の値を求めるために用いる予測結果データ及びピッチ成分データの値の一例を示すグラフである。ただし、Ｙ（ｐ）の値（ただし、ｐは１以上ｎ以下の整数）は、循環シフトを行う前のピッチ成分データのｐ番目のサンプルの値である。従って、例えば、音片データのサンプルが時刻の早い順に並んでおり、循環シフトが下位方向（つまり時刻が遅い方）へと行われるものとすれば、ｊ＜ｐの場合はＹｊ（ｐ）＝Ｙ（ｐ－ｊ）であり、一方、１≦ｐ≦ｊの場合はＹｊ（ｐ）＝Ｙ（ｎ－ｊ＋ｐ）である。

そして、音片編集部８は、上述のＲｘｙ（ｊ）の最大値と、上述の係数ｄｔとに基づいて、定型メッセージ内の音片の読みと一致する音片を表す音片データのうち、数式６の右辺の値（評価値）ｃｏｓｔ２が最大となるものを選択すればよい。

$$cost2 = 1 / (W_3 \mid Rmax \mid + dt)$$

（ただし、W_3は所定の係数、ＲｍａｘはＲｘｙ（０）～Ｒｘｙ（ｎ－１）のうちの最大値）

なお、音片編集部８は、必ずしもピッチ成分データを種々循環シフトしたものについて上述の相関係数を求める必要はなく、例えば、Ｒｘｙ（０）の値をそのまま相関係数の最大値として扱うようにしてもよい。

また、評価値ｃｏｓｔ１やｃｏｓｔ２は、係数ｄｔの項を含まなくてもよく、この場合、音片編集部８は、係数ｄｔを求める必要がない。

あるいは、音片編集部８は、係数ｄｔの値をそのまま評価値として用いてもよく、この場合、音片編集部は、勾配αや、切片βや、

Ｒｘｙ（ｊ）の値を求める必要がない。

また、ピッチ成分データは音片データが表す音片のピッチ長の時間変化を表すデータであってもよい。この場合、音片編集部８は、予測結果データとして、音片のピッチ長の時間変化の予測結果を表すデータを作成するものとし、この音片と読みが合致する音片の波形を表す音片データのピッチ長の時間変化を表すピッチ成分データとの相関を求めるようにすればよい。

また、音片データベース作成部１３は、マイクロフォン、増幅器、サンプリング回路、Ａ／Ｄ（Analog-to-Digital）コンバータ及びＰＣＭエンコーダなどを備えていてもよい。この場合、音片データベース作成部１３は、収録音片データセット記憶部１２より音片データを取得する代わりに、自己のマイクロフォンが集音した音声を表す音声信号を増幅し、サンプリングしてＡ／Ｄ変換した後、サンプリングされた音声信号にＰＣＭ変調を施すことにより、音片データを作成してもよい。

また、音片編集部８は、音響処理部４より返送された波形データを話速変換部１１に供給することにより、当該波形データが表す波形の時間長を、発声スピードデータが示すスピードに合致させるようにしてもよい。

また、音片編集部８は、例えば、言語処理部１と共にフリーテキストデータを取得し、このフリーテキストデータが表すフリーテキストに含まれる音片の波形に最も近い波形を表す音片データを、定型メッセージに含まれる音片の波形に最も近い波形を表す音片データを選択する処理と実質的に同一の処理を行うことによって選択して、音声の合成に用いてもよい。

この場合、音響処理部４は、音片編集部８が選択した音片データが表す音片については、この音片の波形を表す波形データを検索部５に索出させなくてもよい。なお、音片編集部８は、音響処理部４が合成しなくてよい音片を音響処理部４に通知し、音響処理部４はこの通知に応答して、この音片を構成する単位音声の波形の検索を中止するようにすればよい。

また、音片編集部８は、例えば、音響処理部４と共に配信文字列データを取得し、この配信文字列データが表す配信文字列に含まれる音片の波形に最も近い波形を表す音片データを、定型メッセージに含まれる音片の波形に最も近い波形を表す音片データを選択する処理と実質的に同一の処理を行うことによって選択して、音声の合成に用いてもよい。この場合、音響処理部４は、音片編集部８が選択した音片データが表す音片については、この音片の波形を表す波形データを検索部５に索出させなくてもよい。

（第２の実施の形態）

次に、この発明の第２の実施の形態を説明する。この発明の第２の実施の形態に係る音声合成システムの物理的構成は、上述した第１の実施の形態における構成と実質的に同一である。

ただし、第２の実施の形態の音声合成システムにおける音片データベース１０のディレクトリ部ＤＩＲには、例えば第４図に示すように、個々の圧縮音声データについて、上述の（Ａ）～（Ｄ）のデータが互いに対応づけられた形で格納されているほか、上述の（Ｅ）のデータに代え、ピッチ成分データとして、（Ｆ）この圧縮音片データが表す音片の先頭と末尾におけるピッチ成分の周波数を表すデータが、これら（Ａ）～（Ｄ）のデータに対応付けられた形で格納

されている。

なお、第４図は、第２図と同様、データ部ＤＡＴに含まれるデータとして、読みが「サイタマ」である音片の波形を表す、データ量１４１０ｈバイトの圧縮音片データが、アドレス００１Ａ３６Ａ６ｈを先頭とする論理的位置に格納されている場合を例示している。）また、上述の（Ａ）～（Ｄ）及び（Ｆ）のデータの集合のうち少なくとも（Ａ）のデータは、音片読みデータが表す表音文字に基づいて決められた順位に従ってソートされた状態で音片データベース１０の記憶領域に格納されているものとする。

そして、音片登録ユニットＲの音片データベース作成部１３は、収録音片データセット記憶部１２より、互いに対応付けられている表音文字及び音片データを読み出すと、この音片データが表す音声の発声スピードと、先頭及び末尾でのピッチ成分の周波数とを特定するものとする。

そして、読み出した音片データを圧縮部１４に供給し、圧縮音片データの返送を受けると、この圧縮音片データ、収録音片データセット記憶部１２より読み出した表音文字、この圧縮音片データの音片データベース１０の記憶領域内での先頭のアドレス、この圧縮音片データのデータ長、及び、特定した発声スピードを示すスピード初期値データを、第１の実施の形態の音片データベース作成部１３と同様の動作を行うことにより音片データベース１０の記憶領域に書き込み、また、音声の先頭及び末尾におけるピッチ成分の周波数を特定した結果を示すデータを生成して、ピッチ成分データとして音片データベース１０の記憶領域に書き込むものとする。

なお、発声スピード及びピッチ成分の周波数の特定は、例えば、

第１の実施の形態の音片データベース作成部１３が行う手法と実質的に同一の手法により行えばよい。

次に、この音声合成システムの動作を説明する。

この音声合成システムの言語処理部１がフリーテキストデータを外部から取得した場合、及び、音響処理部４が配信文字列データを取得した場合の動作は、第１の実施の形態の音声合成システムが行う動作と実質的に同一である。（なお、言語処理部１がフリーテキストデータを取得する手法や音響処理部４が配信文字列データを取得する手法はいずれも任意であり、例えば、いずれも第１の実施の形態における言語処理部１や音響処理部４が行う手法と同様の手法によりフリーテキストデータあるいは配信文字列データを取得すればよい。）

次に、音片編集部８が、定型メッセージデータ及び発声スピードデータを取得したとする。（なお、音片編集部８が定型メッセージデータや発声スピードデータを取得する手法も任意であり、例えば、第１の実施の形態の音片編集部８が行う手法と同様の手法で定型メッセージデータや発声スピードデータを取得すればよい。）

定型メッセージデータ及び発声スピードデータが音片編集部８に供給されると、音片編集部８は、第１の実施の形態における音片編集部８と同様に、定型メッセージに含まれる音片の読みを表す表音文字に合致する表音文字が対応付けられている圧縮音片データをすべて索出するよう、検索部９に指示する。また、話速変換部１１に対しても、第１の実施の形態における音片編集部８と同様に、話速変換部１１に供給される音片データを変換して、当該音片データが表す音片の時間長を、発声スピードデータが示すスピードに合致す

るようにすることを指示する。

すると、検索部９、伸張部６及び話速変換部１１が、第１の実施の形態における検索部９、伸張部６及び話速変換部１１の動作と実質的に同一の動作を行い、この結果、話速変換部１１から音片編集部８へと、音片データ、音片読みデータ及びピッチ成分データが供給される。また、欠落部分識別データが検索部９より話速変換部１１へと供給された場合は、更にこの欠落部分識別データも音片編集部８へと供給される。

音片編集部８は、話速変換部１１より音片データ、音片読みデータ及びピッチ成分データを供給されると、以下説明する手順に従い、供給された音片データのうちから、定型メッセージを構成する音片の波形とみなせる波形を表す音片データを、音片１個につき１個ずつ選択する。

具体的には、まず、音片編集部８は、話速変換部１１より供給されたピッチ成分データに基づき、話速変換部１１より供給された各音片データの先頭及び末尾の各時点でのピッチ成分の周波数を特定する。そして、話速変換部１１より供給された音片データのうちから、定型メッセージ内で隣接する音片同士の境界でのピッチ成分の周波数の差の絶対値を定型メッセージ全体で累計した値が最小になる、という条件を満たすように、音片データを選択する。

音片データを選択する条件を、第５図（ａ）～（ｄ）を参照して説明する。例えば、第５図（ａ）に示すような、「このさきみぎかーぶです」という読みの定型メッセージを表す定型メッセージデータが音片編集部８に供給されたものとし、この定型メッセージが「このさき」、「みぎかーぶ」及び「です」という３個の音片からなる

ものとする。そして、第５図（ｂ）にリストを示すように、音片データベース１０が、読みが「このさき」である圧縮音片データが３個（第５図（ｂ）において「Ａ１」「Ａ２」あるいは「Ａ３」として表したもの）、読みが「みぎかーぶ」である圧縮音片データが２個（第５図（ｂ）において「Ｂ１」あるいは「Ｂ２」として表したもの）、読みが「です」である圧縮音片データが３個（第５図（ｂ）において「Ｃ１」「Ｃ２」あるいは「Ｃ３」として表したもの）、それぞれ索出され、伸長され、音片データとして音片編集部８へと供給されたとする。

一方、読みが「このさき」である各音片データが表す各音片の末尾におけるピッチ成分の周波数と読みが「みぎかーぶ」である各音片データが表す各音片の先頭におけるピッチ成分の周波数との差の絶対値は第５図（ｃ）に示す通りであったとする。（第５図（ｃ）は、例えば、音片データＡ１が表す音片の末尾におけるピッチ成分の周波数と音片データＢ１が表す音片の先頭におけるピッチ成分の周波数との差の絶対値は「１２３」であることを示している。なお、この絶対値の単位は、例えば「ヘルツ」である。）

また、読みが「みぎかーぶ」である各音片データが表す各音片の末尾におけるピッチ成分の周波数と読みが「です」である各音片データが表す各音片の先頭におけるピッチ成分の周波数との差の絶対値は第５図（ｃ）に示す通りであったとする。

この場合において、「このさきみぎかーぶです」という定型メッセージを読み上げる音声の波形を音片データを用いて生成した場合、隣接する音片同士の境界でのピッチ成分の周波数の差の絶対値の累計が最小になる組み合わせは、Ａ３、Ｂ２及びＣ２という組み合わ

せである。従ってこの場合、音片編集部８は、第５図（ｄ）に示すように、音片データＡ３、Ｂ２及びＣ２を選択する。

この条件を満たす音片データを選択するために、音片編集部８は、例えば、定型メッセージ内で隣接する音片同士の境界でのピッチ成分の周波数の差の絶対値を距離として定義し、ＤＰ（Dynamic Programming）マッチングの手法により音片データを選ぶようにすればよい。

一方、音片編集部８は、話速変換部１１より欠落部分識別データも供給されている場合には、欠落部分識別データが示す音片の読みを表す表音文字列を定型メッセージデータより抽出して音響処理部４に供給し、この音片の波形を合成するよう指示する。

指示を受けた音響処理部４は、音片編集部８より供給された表音文字列を、配信文字列データが表す表音文字列と同様に扱う。この結果、この表音文字列に含まれる表音文字が示す音声の波形を表す圧縮波形データが検索部５により索出され、この圧縮波形データが伸長部６により元の波形データへと復元され、検索部５を介して音響処理部４へと供給される。音響処理部４は、この波形データを音片編集部８へと供給する。

音片編集部８は、音響処理部４より波形データを返送されると、この波形データと、話速変換部１１より供給された音片データのうち音片編集部８が選択したものとを、定型メッセージデータが示す定型メッセージ内での各音片の並びに従った順序で互いに結合し、合成音声を表すデータとして出力する。

なお、話速変換部１１より供給されたデータに欠落部分識別データが含まれていない場合は、第１の実施の形態と同様、音響処理部

４に波形の合成を指示することなく直ちに、音片編集部８が選択した音片データを、定型メッセージデータが示す定型メッセージ内での各音片の並びに従った順序で互いに結合し、合成音声を表すデータとして出力すればよい。

以上説明したように、この第２の実施の形態の音声合成システムでは、音片データ同士の境界でのピッチ成分の周波数の不連続的な変化の量の累計が定型メッセージ全体で最小となるように音片データが選ばれ、録音編集方式により自然につなぎ合わせられるため、合成音声が自然なものとなる。また、この音声合成システムでは、処理が複雑な韻律予測は行われないので、簡単な構成で高速な処理にも追随できる。

なお、この第２の実施の形態の音声合成システムの構成も、上述のものに限られない。

例えば、ピッチ成分データは音片データが表す音片の先頭及び末尾でのピッチ長を表すデータであってもよい。この場合、音片編集部８は、話速変換部１１より供給された各音片データの先頭及び末尾でのピッチ長を話速変換部１１より供給されたピッチ成分データに基づいて特定し、定型メッセージ内で隣接する音片同士の境界でのピッチ長の差の絶対値を定型メッセージ全体で累計した値が最小になる、という条件を満たすように、音片データを選択すればよい。

また、音片編集部８は、例えば、言語処理部１と共にフリーテキストデータを取得し、このフリーテキストデータが表すフリーテキストに含まれる音片の波形とみなせる波形を表す音片データを、定型メッセージに含まれる音片の波形とみなせる波形を表す音片データを抽出する処理と実質的に同一の処理を行うことによって抽出し

て、音声の合成に用いてもよい。

この場合、音響処理部４は、音片編集部８が抽出した音片データが表す音片については、この音片の波形を表す波形データを検索部５に索出させなくてもよい。なお、音片編集部８は、音響処理部４が合成しなくてよい音片を音響処理部４に通知し、音響処理部４はこの通知に応答して、この音片を構成する単位音声の波形の検索を中止するようにすればよい。

また、音片編集部８は、例えば、音響処理部４と共に配信文字列データを取得し、この配信文字列データが表す配信文字列に含まれる音片の波形とみなせる波形を表す音片データを、定型メッセージに含まれる音片の波形とみなせる波形を表す音片データを抽出する処理と実質的に同一の処理を行うことによって抽出して、音声の合成に用いてもよい。この場合、音響処理部４は、音片編集部８が抽出した音片データが表す音片については、この音片の波形を表す波形データを検索部５に索出させなくてもよい。

（第３の実施の形態）

次に、この発明の第３の実施の形態を説明する。この発明の第３の実施の形態に係る音声合成システムの物理的構成は、上述した第１の実施の形態における構成と実質的に同一である。

次に、この音声合成システムの動作を説明する。

この音声合成システムの言語処理部１がフリーテキストデータを外部から取得した場合、及び、音響処理部４が配信文字列データを取得した場合の動作は、第１又は第２の実施の形態の音声合成システムが行う動作と実質的に同一である。（なお、言語処理部１がフリーテキストデータを取得する手法や音響処理部４が配信文字列デ

ータを取得する手法はいずれも任意であり、例えば、いずれも第１又は第２の実施の形態における言語処理部１や音響処理部４が行う手法と同様の手法によりフリーテキストデータあるいは配信文字列データを取得すればよい。）

次に、音片編集部８が、定型メッセージデータ及び発声スピードデータを取得したとする。なお、音片編集部８が定型メッセージデータや発声スピードデータを取得する手法も任意であり、例えば、第１の実施の形態の音片編集部８が行う手法と同様の手法で定型メッセージデータや発声スピードデータを取得すればよい。あるいは、例えばこの音声合成システムがカーナビゲーションシステム等の車両内システムの一部をなすものであって、この車両内システムを構成するの他の装置（例えば、音声認識を行い、音声認識の結果得られた情報に基づいてエージェント処理を実行する装置など）が、ユーザーに対して発話する内容や発話スピードを決定し、決定結果を表すデータを生成するものである場合、この音声合成システムは、生成されたこのデータを受信（取得）し、定型メッセージデータ及び発声スピードデータとして扱うようにしてもよい。

定型メッセージデータ及び発声スピードデータが音片編集部８に供給されると、音片編集部８は、第１の実施の形態における音片編集部８と同様に、定型メッセージに含まれる音片の読みを表す表音文字に合致する表音文字が対応付けられている圧縮音片データをすべて索出するよう、検索部９に指示する。また、話速変換部１１に対しても、第１の実施の形態における音片編集部８と同様に、話速変換部１１に供給される音片データを変換して、当該音片データが表す音片の時間長を、発声スピードデータが示すスピードに合致す

るようにすることを指示する。

すると、検索部９、伸張部６及び話速変換部１１が、第１の実施の形態における検索部９、伸張部６及び話速変換部１１の動作と実質的に同一の動作を行い、この結果、話速変換部１１から音片編集部８へと、音片データ、音片読みデータ、この音片データが表す音片の発声スピードを表すスピード初期値データ及びピッチ成分データが供給される。また、欠落部分識別データが検索部９より話速変換部１１へと供給された場合は、更にこの欠落部分識別データも音片編集部８へと供給される。

音片編集部８は、話速変換部１１より音片データ、音片読みデータ及びピッチ成分データを供給されると、話速変換部１１より供給された各々のピッチ成分データについて上述の値α、βの組及び／又はＲｍａｘを求め、また、このスピード初期値データと、音片編集部８に供給された定型メッセージデータ及び発声スピードデータとを用いて、上述の値ｄｔを求める。

そして、音片編集部８は、話速変換部１１より供給されたそれぞれの音片データにつき、自ら求めた当該音片データ（以下、音片データＸと記す）についてのα、β、Ｒｍａｘ及びｄｔの値と、定型メッセージ内で当該音片データが表す音片の後に隣接する音片を表す音片データ（以下、音片データＹと記す）のピッチ成分の周波数とに基づいて、数式７に示す評価値H_{XY}を特定する。

$H_{XY}=(W_A\cdot cost\_A)+(W_B\cdot cost\_B)+(W_C\cdot cost\_C)$

（ただし、W_A、W_B及びW_Cはいずれも所定の係数であり、W_Aは０ではないものとする）

数式７の右辺に含まれる値ｃｏｓｔ＿Ａは、当該定型メッセージ内で互いに隣接する、音片データＸが表す音片と音片データＹが表す音片との境界でのピッチ成分の周波数の差の絶対値の逆数である。

なお、音片編集部８は、ｃｏｓｔ＿Ａの値を特定するため、話速変換部１１より供給されたピッチ成分データに基づき、話速変換部１１より供給された各音片データの先頭及び末尾の各時点でのピッチ成分の周波数を特定するようにすればよい。

また、数式７の右辺に含まれる値ｃｏｓｔ＿Ｂは、音片データＸについて数式８に従って評価値ｃｏｓｔ＿Ｂを求めた場合の値である。

$$cost\_B = 1 / (W_{B1} \mid 1-\alpha \mid + W_{B2} \mid \beta \mid + W_{B3} \cdot dt)$$

（ただし、W_{B1}、W_{B2}及びW_{B3}は所定の正の係数）

また、数式７の右辺に含まれる値ｃｏｓｔ＿Ｃは、音片データＸについて数式９に従って評価値ｃｏｓｔ＿Ｃを求めた場合における値である。

$$cost\_C = 1 / (W_{C1} \mid Rmax \mid + W_{C2} \cdot dt)$$

（ただし、W_{C1}及びW_{C2}は所定の係数）

あるいは、音片編集部８は、数式７～数式９に代えて、数式１０及び数式１１に従って評価値H_{XY}を特定するようにしてもよい。ただし、数式１０に含まれるｃｏｓｔ＿Ｂ及びｃｏｓｔ＿Ｃについては、上述の係数W_{B3}及びW_{C3}の値はいずれも０とする。また、数式８及び数式９における（$W_{B3}\cdot dt$）及び（$W_{C2}\cdot dt$）の項を備えなくともよい。

$$H_{XY} = (W_A \cdot cost\_A) + (W_B \cdot cost\_B) + (W_C \cdot$$

$cost\_C) + (W_D \cdot cost\_D)$

（ただし、W_Dは０でない所定の係数）

$cost\_D = 1 / (W_{d1} \cdot dt)$

（ただし、W_{d1}は０でない所定の係数）

そして、音片編集部８は、話速変換部１１より供給された各音片データのうちから、音片編集部８に供給された定型メッセージデータが表す定型メッセージを構成する音片１個につき１個ずつの音片データを選ぶことにより得られる各組み合わせのうち、組み合わせに属する各音片データの評価値H_{XY}の総和が最大となるものを、定型メッセージを読み上げる音声を合成するための最適な音片データの組み合わせとして選択する。

つまり、例えば第５図に示すように、定型メッセージデータが表す定型メッセージが音片Ａ，Ｂ及びＣより構成され、音片Ａを表す音片データの候補として音片データＡ１，Ａ２及びＡ３が索出され、音片Ｂを表す音片データの候補として音片データＢ１及びＢ２が索出され、音片Ｃを表す音片データの候補として音片データＣ１，Ｃ２及びＣ３が索出された場合、音片データＡ１，Ａ２及びＡ３のうちから１個、音片データＢ１及びＢ２のうちから１個、音片データＣ１，Ｃ２及びＣ３のうちから１個、計３個選ぶことにより得られる組み合わせ計１８通りのうち、組み合わせに属する各音片データの評価値H_{XY}の総和が最大となるものを、定型メッセージを読み上げる音声を合成するための最適な音片データの組み合わせとして選択する。

ただし、総和を求めるために用いられる評価値H_{XY}としては、組み合わせ内での音片の接続関係を正しく反映したものが選ばれるも

のとする。つまり、例えば組み合わせ内に、音片ｐを表す音片データＰ及び音片ｑを表す音片データＱが含まれており、定型メッセージ内では音片ｐが音片ｑに先行する形で互いに隣接するという場合、音片データＰの評価値としては、音片ｐが音片ｑに先行する形で互いに隣接する場合における評価値H_{PQ}が用いられるものとする。

また、定型メッセージの末尾の音片（例えば、第５図を参照して前述した例でいえば、音片Ｃ１，Ｃ２及びＣ３）については、後続する音片が存在しないため、ｃｏｓｔ＿Ａの値を定めることができない。このため、これら末尾の音片を表す音片データの評価値H_{XY}を算定するにあたって、音片編集部８は、（W_A・ｃｏｓｔ＿Ａ）の値を０であるものとして扱い、一方、係数W_B，W_C及びW_Dの値は、それぞれ、他の音片データの評価値H_{XY}を算定する場合とは異なる所定の値であるものとして扱う。

なお、音片編集部８は、数式７あるいは数式１１を用いて、音片データＸについて、当該音片データＸが表す音片の前に隣接する音片データＹとの関係を表す評価値を含むものとして評価値H_{XY}を特定してもよい。この場合は、定型メッセージの先頭の音片について、先行する音片が存在しないため、ｃｏｓｔ＿Ａの値を定めることができないこととなる。このため、これら先頭の音片を表す音片データの評価値H_{XY}を算定するにあたって、音片編集部８は、（W_A・ｃｏｓｔ＿Ａ）の値を０であるものとして扱い、一方、係数W_B，W_C及びW_Dの値は、それぞれ、他の音片データの評価値H_{XY}を算定する場合とは異なる所定の値であるものとして扱うようにすればよい。

一方、音片編集部８は、話速変換部１１より欠落部分識別データ

も供給されている場合には、欠落部分識別データが示す音片の読みを表す表音文字列を定型メッセージデータより抽出して音響処理部４に供給し、この音片の波形を合成するよう指示する。

指示を受けた音響処理部４は、音片編集部８より供給された表音文字列を、配信文字列データが表す表音文字列と同様に扱う。この結果、この表音文字列に含まれる表音文字が示す音声の波形を表す圧縮波形データが検索部５により索出され、この圧縮波形データが伸長部６により元の波形データへと復元され、検索部５を介して音響処理部４へと供給される。音響処理部４は、この波形データを音片編集部８へと供給する。

音片編集部８は、音響処理部４より波形データを返送されると、この波形データと、話速変換部１１より供給された音片データのうち、評価値H_{XY}の総和が最大となる組み合わせとして音片編集部８が選択した組み合わせに属するものとを、定型メッセージデータが示す定型メッセージ内での各音片の並びに従った順序で互いに結合し、合成音声を表すデータとして出力する。

なお、話速変換部１１より供給されたデータに欠落部分識別データが含まれていない場合は、第１の実施の形態と同様、音響処理部４に波形の合成を指示することなく直ちに、音片編集部８が選択した音片データを、定型メッセージデータが示す定型メッセージ内での各音片の並びに従った順序で互いに結合し、合成音声を表すデータとして出力すればよい。

以上説明したように、この第３の実施の形態の音声合成システムでも、音片データが録音編集方式により自然につなぎ合わせられ、定型メッセージを読み上げる音声が合成される。音片データベース

１０の記憶容量は、音素毎に波形を記憶する場合に比べて小さくでき、また、高速に検索できる。このため、この音声合成システムは小型軽量に構成することができ、また高速な処理にも追随できる。

そして、第３の実施の形態の音声合成システムによれば、定型メッセージを読み上げる音声を合成するために選択される音片データの組み合わせの適切さを評価するための様々な評価基準（例えば、音片の波形の予測結果と音片データとの相関を１次回帰させた場合の勾配や切片による評価や、音片の時間差による評価や、音片データ同士の境界でのピッチ成分の周波数の不連続的な変化の量の累計、など）が、１個の評価値に影響を及ぼす形で総合的に反映され、この結果、最も自然な合成音声を合成するために選択すべき最適な音片データの組み合わせが、適正に決定される。

なお、この第３の実施の形態の音声合成システムの構成も、上述のものに限られない。

例えば、最適な音片データの組み合わせを選択するために音片編集部８が用いる評価値は数式７～１３に示すものに限られず、音片データが表す音片を互いに結合して得られる音声が、人の発する音声にどの程度類似又は相違しているかについての評価を表す任意の値であってよい。

また、評価値を表す数式（評価式）に含まれる変数ないし定数も必ずしも数式７～１３に含まれているものに限られず、評価式としては、音片データが表す音片の特徴を示す任意のパラメータや、あるいは当該音片を互いに結合して得られる音声の特徴を示す任意のパラメータや、あるいは当該音声を人が発した場合に当該音声に備わると予測される特徴を示す任意のパラメータを含んだ数式が用い

られてよい。

また、最適な音片データの組み合わせを選択するための基準は必ずしも評価値の形で表現可能なものである必要はなく、音片データが表す音片を互いに結合して得られる音声が人の発する音声にどの程度類似又は相違しているかについての評価に基づいて音片データの最適な組み合わせを特定するに至るような基準である限り任意である。

また、音片編集部８は、例えば、言語処理部１と共にフリーテキストデータを取得し、このフリーテキストデータが表すフリーテキストに含まれる音片の波形とみなせる波形を表す音片データを、定型メッセージに含まれる音片の波形とみなせる波形を表す音片データを抽出する処理と実質的に同一の処理を行うことによって抽出して、音声の合成に用いてもよい。この場合、音響処理部４は、音片編集部８が抽出した音片データが表す音片については、この音片の波形を表す波形データを検索部５に索出させなくてもよい。なお、音片編集部８は、音響処理部４が合成しなくてよい音片を音響処理部４に通知し、音響処理部４はこの通知に応答して、この音片を構成する単位音声の波形の検索を中止するようにすればよい。

また、音片編集部８は、例えば、音響処理部４と共に配信文字列データを取得し、この配信文字列データが表す配信文字列に含まれる音片の波形とみなせる波形を表す音片データを、定型メッセージに含まれる音片の波形とみなせる波形を表す音片データを抽出する処理と実質的に同一の処理を行うことによって抽出して、音声の合成に用いてもよい。この場合、音響処理部４は、音片編集部８が抽出した音片データが表す音片については、この音片の波形を表す波

形データを検索部５に索出させなくてもよい。

以上、この発明の実施の形態を説明したが、この発明にかかる音声データ選択装置は、専用のシステムによらず、通常のコンピュータシステムを用いて実現可能である。

例えば、パーソナルコンピュータに上述の第１の実施の形態における言語処理部１、一般単語辞書２、ユーザ単語辞書３、音響処理部４、検索部５、伸長部６、波形データベース７、音片編集部８、検索部９、音片データベース１０及び話速変換部１１の動作を実行させるためのプログラムを格納した媒体（ＣＤ－ＲＯＭ、ＭＯ、フロッピー（登録商標）ディスク等）から該プログラムをインストールすることにより、当該パーソナルコンピュータに、上述の第１の実施の形態の本体ユニットＭの機能を行わせることができる。

また、パーソナルコンピュータに、上述の第１の実施の形態における収録音片データセット記憶部１２、音片データベース作成部１３及び圧縮部１４の動作を実行させるためのプログラムを格納した媒体から該プログラムをインストールすることにより、当該パーソナルコンピュータに、上述の第１の実施の形態の音片登録ユニットＲの機能を行わせることができる。

そして、これらのプログラムを実行し、第１の実施の形態における本体ユニットＭや音片登録ユニットＲとして機能するパーソナルコンピュータが、第１図の音声合成システムの動作に相当する処理として、第６図～第８図に示す処理を行うものとする。

第６図は、このパーソナルコンピュータがフリーテキストデータを取得した場合の処理を示すフローチャートである。

第７図は、このパーソナルコンピュータが配信文字列データを取

得した場合の処理を示すフローチャートである。

第８図は、このパーソナルコンピュータが定型メッセージデータ及び発声スピードデータを取得した場合の処理を示すフローチャートである。

すなわち、まず、このパーソナルコンピュータが、外部より、上述のフリーテキストデータを取得すると（第６図、ステップＳ１０１）、このフリーテキストデータが表すフリーテキストに含まれるそれぞれの表意文字について、その読みを表す表音文字を、一般単語辞書２やユーザ単語辞書３を検索することにより特定し、この表意文字を、特定した表音文字へと置換する（ステップＳ１０２）。なお、このパーソナルコンピュータがフリーテキストデータを取得する手法は任意である。

そして、このパーソナルコンピュータは、フリーテキスト内の表意文字をすべて表音文字へと置換した結果を表す表音文字列が得られると、この表音文字列に含まれるそれぞれの表音文字について、当該表音文字が表す単位音声の波形を波形データベース７より検索し、表音文字列に含まれるそれぞれの表音文字が表す単位音声の波形を表す圧縮波形データを索出する（ステップＳ１０３）。

次に、このパーソナルコンピュータは、索出された圧縮波形データを、圧縮される前の波形データへと復元し（ステップＳ１０４）、復元された波形データを、表音文字列内での各表音文字の並びに従った順序で互いに結合し、合成音声データとして出力する（ステップＳ１０５）。なお、このパーソナルコンピュータが合成音声データを出力する手法は任意である。

また、このパーソナルコンピュータが、外部より、上述の配信文

字列データを任意の手法で取得すると(第７図､ステップＳ２０１)、この配信文字列データが表す表音文字列に含まれるそれぞれの表音文字について、当該表音文字が表す単位音声の波形を波形データベース７より検索し、表音文字列に含まれるそれぞれの表音文字が表す単位音声の波形を表す圧縮波形データを索出する（ステップＳ２０２）。

次に、このパーソナルコンピュータは、索出された圧縮波形データを､圧縮される前の波形データへと復元し(ステップＳ２０３)、復元された波形データを、表音文字列内での各表音文字の並びに従った順序で互いに結合し、合成音声データとしてステップＳ１０５の処理と同様の処理により出力する（ステップＳ２０４）。

一方、このパーソナルコンピュータが、外部より、上述の定型メッセージデータ及び発声スピードデータを任意の手法により取得すると（第８図、ステップＳ３０１）、まず、この定型メッセージデータが表す定型メッセージに含まれる音片の読みを表す表音文字に合致する表音文字が対応付けられている圧縮音片データをすべて索出する（ステップＳ３０２）。

また、ステップＳ３０２では、該当する圧縮音片データに対応付けられている上述の音片読みデータ、スピード初期値データ及びピッチ成分データも索出する。なお、１個の音片につき複数の圧縮音片データが該当する場合は、該当する圧縮音片データすべてを索出する。一方、圧縮音片データを索出できなかった音片があった場合は、上述の欠落部分識別データを生成する。

次に、このパーソナルコンピュータは、索出された圧縮音片データを､圧縮される前の音片データへと復元する(ステップＳ３０３)。

そして、復元された音片データを、上述の音片編集部８が行う処理と同様の処理により変換して、当該音片データが表す音片の時間長を、発声スピードデータが示すスピードに合致させる（ステップＳ３０４）。なお、発声スピードデータが供給されていない場合は、復元された音片データを変換しなくてもよい。

次に、このパーソナルコンピュータは、音片の時間長が変換された音片データのうちから、定型メッセージを構成する音片の波形に最も近い波形を表す音片データを、上述の音片編集部８が行う処理と同様の処理を行うことにより、音片１個につき１個ずつ選択する（ステップＳ３０５～Ｓ３０８）。

すなわち、このパーソナルコンピュータは、定型メッセージデータが表す定型メッセージに韻律予測の手法に基づいた解析を加えることにより、この定型メッセージの韻律を予測する（ステップＳ３０５）。そして、定型メッセージ内のそれぞれの音片について、この音片のピッチ成分の周波数の時間変化の予測結果と、この音片と読みが合致する音片の波形を表す音片データのピッチ成分の周波数の時間変化を表すピッチ成分データとの相関を求める（ステップＳ３０６）。より具体的には、索出された各々のピッチ成分データについて、例えば、上述した勾配α及び切片βの値を求める。

一方で、このパーソナルコンピュータは、索出されたスピード初期値データと、外部より取得した定型メッセージデータ及び発声スピードデータとを用いて、上述の値ｄｔを求める（ステップＳ３０７）。

そして、このパーソナルコンピュータは、ステップＳ３０６で求めたα、βの値、及び、ステップＳ３０７で求めたｄｔの値に基づ

いて、定型メッセージ内の音片の読みと一致する音片を表す音片データのうち、上述の評価値ｃｏｓｔ１が最大となるものを選択する（ステップＳ３０８）。

なお、このパーソナルコンピュータは、ステップＳ３０６で、上述のα及びβの値を求める代わりに、上述のＲｘｙ（ｊ）の最大値を求めるようにしてもよい。この場合は、ステップＳ３０８で、Ｒｘｙ（ｊ）の最大値と、ステップＳ３０７で求めた係数ｄｔとに基づいて、定型メッセージ内の音片の読みと一致する音片を表す音片データのうち、上述の評価値ｃｏｓｔ２が最大となるものを選択すればよい。

一方、このパーソナルコンピュータは、欠落部分識別データを生成した場合、欠落部分識別データが示す音片の読みを表す表音文字列を定型メッセージデータより抽出し、この表音文字列につき、音素毎に、配信文字列データが表す表音文字列と同様に扱って上述のステップＳ２０２～Ｓ２０３の処理を行うことにより、この表音文字列内の各表音文字が示す音声の波形を表す波形データを復元する（ステップＳ３０９）。

そして、このパーソナルコンピュータは、復元した波形データと、ステップＳ３０８で選択した音片データとを、定型メッセージデータが示す定型メッセージ内での各音片の並びに従った順序で互いに結合し、合成音声を表すデータとして出力する（ステップＳ３１０）。

また、パーソナルコンピュータに上述の第２の実施の形態における言語処理部１、一般単語辞書２、ユーザ単語辞書３、音響処理部４、検索部５、伸長部６、波形データベース７、音片編集部８、検索部９、音片データベース１０及び話速変換部１１の動作を実行さ

せるためのプログラムを格納した媒体から該プログラムをインストールすることにより、当該パーソナルコンピュータに、上述の第２の実施の形態における本体ユニットＭの機能を行わせることができる。

また、パーソナルコンピュータに上述の第２の実施の形態における収録音片データセット記憶部１２、音片データベース作成部１３及び圧縮部１４の動作を実行させるためのプログラムを格納した媒体から該プログラムをインストールすることにより、当該パーソナルコンピュータに、上述の第２の実施の形態における音片登録ユニットＲの機能を行わせることができる。

そして、これらのプログラムを実行し、第２の実施の形態における本体ユニットＭや音片登録ユニットＲとして機能するパーソナルコンピュータが、第１図の音声合成システムの動作に相当する処理として、第６図及び第７図に示す上述の処理を行い、また、第９図に示す処理を行うものとする。

第９図は、このパーソナルコンピュータが定型メッセージデータ及び発声スピードデータを取得した場合の処理を示すフローチャートである。

すなわち、このパーソナルコンピュータが、外部より、上述の定型メッセージデータ及び発声スピードデータを任意の手法により取得すると（第９図、ステップＳ４０１）、まず、上述のステップＳ３０２の処理と同様に、この定型メッセージデータが表す定型メッセージに含まれる音片の読みを表す表音文字に合致する表音文字が対応付けられている圧縮音片データと、該当する圧縮音片データに対応付けられている上述の音片読みデータ、スピード初期値データ

及びピッチ成分データとを、すべて索出する（ステップＳ４０２）。なお、ステップＳ４０２でも、１個の音片につき複数の圧縮音片データが該当する場合は該当する圧縮音片データすべてを索出し、一方で圧縮音片データを索出できなかった音片があった場合は、上述の欠落部分識別データを生成する。

次に、このパーソナルコンピュータは、索出された圧縮音片データを、圧縮される前の音片データへと復元し（ステップＳ４０３）、復元された音片データを、上述の音片編集部８が行う処理と同様の処理により変換して、当該音片データが表す音片の時間長を、発声スピードデータが示すスピードに合致させる（ステップＳ４０４）。なお、発声スピードデータが供給されていない場合は、復元された音片データを変換しなくてもよい。

次に、このパーソナルコンピュータは、音片の時間長が変換された音片データのうちから、定型メッセージを構成する音片の波形とみなせる波形を表す音片データを、上述の第２の実施の形態における音片編集部８が行う処理と同様の処理を行うことにより、音片１個につき１個ずつ選択する（ステップＳ４０５～Ｓ４０６）。

具体的には、まず、このパーソナルコンピュータは、音片の時間長が変換された各音片データの先頭及び末尾の各時点でのピッチ成分の周波数を、索出されたピッチ成分データに基づいて特定する（ステップＳ４０５）。そして、これらの音片データのうちから、定型メッセージ内で隣接する音片同士の境界でのピッチ成分の周波数の差の絶対値を定型メッセージ全体で累計した値が最小になる、という条件を満たすように、音片データを選択する（ステップＳ４０６）。この条件を満たす音片データを選択するために、このパーソナルコ

ンピュータは、例えば、定型メッセージ内で隣接する音片同士の境界でのピッチ成分の周波数の差の絶対値を距離として定義し、ＤＰマッチングの手法により音片データを選ぶようにすればよい。

一方、このパーソナルコンピュータは、欠落部分識別データを生成した場合、欠落部分識別データが示す音片の読みを表す表音文字列を定型メッセージデータより抽出し、この表音文字列につき、音素毎に、配信文字列データが表す表音文字列と同様に扱って上述のステップＳ２０２～Ｓ２０３の処理を行うことにより、この表音文字列内の各表音文字が示す音声の波形を表す波形データを復元する（ステップＳ４０７）。

そして、このパーソナルコンピュータは、復元した波形データと、ステップＳ４０６で選択した音片データとを、定型メッセージデータが示す定型メッセージ内での各音片の並びに従った順序で互いに結合し、合成音声を表すデータとして出力する（ステップＳ４０８）。

また、パーソナルコンピュータに上述の第３の実施の形態における言語処理部１、一般単語辞書２、ユーザ単語辞書３、音響処理部４、検索部５、伸長部６、波形データベース７、音片編集部８、検索部９、音片データベース１０及び話速変換部１１の動作を実行させるためのプログラムを格納した媒体から該プログラムをインストールすることにより、当該パーソナルコンピュータに、上述の第３の実施の形態における本体ユニットＭの機能を行わせることができる。

また、パーソナルコンピュータに上述の第３の実施の形態における収録音片データセット記憶部１２、音片データベース作成部１３及び圧縮部１４の動作を実行させるためのプログラムを格納した媒

体から該プログラムをインストールすることにより、当該パーソナルコンピュータに、上述の第３の実施の形態における音片登録ユニットＲの機能を行わせることができる。

そして、これらのプログラムを実行し、第３の実施の形態における本体ユニットＭや音片登録ユニットＲとして機能するパーソナルコンピュータが、第１図の音声合成システムの動作に相当する処理として、第６図及び第７図に示す上述の処理を行い、また、第１０図に示す処理を行うものとする。

第１０図は、このパーソナルコンピュータが定型メッセージデータ及び発声スピードデータを取得した場合の処理を示すフローチャートである。

すなわち、このパーソナルコンピュータが、外部より、上述の定型メッセージデータ及び発声スピードデータを任意の手法により取得すると（第１０図、ステップＳ５０１）、まず、上述のステップＳ３０２の処理と同様に、この定型メッセージデータが表す定型メッセージに含まれる音片の読みを表す表音文字に合致する表音文字が対応付けられている圧縮音片データと、該当する圧縮音片データに対応付けられている上述の音片読みデータ、スピード初期値データ及びピッチ成分データとを、すべて索出する（ステップＳ５０２）。なお、ステップＳ５０２でも、１個の音片につき複数の圧縮音片データが該当する場合は該当する圧縮音片データすべてを索出し、一方で圧縮音片データを索出できなかった音片があった場合は、上述の欠落部分識別データを生成する。

次に、このパーソナルコンピュータは、索出された圧縮音片データを、圧縮される前の音片データへと復元し（ステップＳ５０３）、

復元された音片データを、上述の音片編集部８が行う処理と同様の処理により変換して、当該音片データが表す音片の時間長を、発声スピードデータが示すスピードに合致させる(ステップＳ５０４)。なお、発声スピードデータが供給されていない場合は、復元された音片データを変換しなくてもよい。

次に、このパーソナルコンピュータは、音片の時間長が変換された音片データのうちから、定型メッセージを読み上げる音声を合成するための最適な音片データの組み合わせを、上述の第３の実施の形態における音片編集部８が行う処理と同様の処理を行うことにより選択する（ステップＳ５０５～Ｓ５０７）。

すなわち、まず、このパーソナルコンピュータは、ステップＳ５０２で索出された各々のピッチ成分データについて上述の値α、βの組及び／又はＲｍａｘを求め、また、このスピード初期値データと、ステップＳ５０１で取得した定型メッセージデータ及び発声スピードデータとを用いて、上述の値ｄｔを求める（ステップＳ５０５）。

次に、このパーソナルコンピュータは、ステップＳ５０４で変換されたそれぞれの音片データにつき、ステップＳ５０５で求めたα、β、Ｒｍａｘ及びｄｔの値と、定型メッセージ内で当該音片データが表す音片の後に隣接する音片を表す音片データのピッチ成分の周波数とに基づいて、上述した評価値H_{XY}を特定する（ステップＳ５０６）。

そして、このパーソナルンピュータは、ステップＳ５０４で変換された各音片データのうちから、ステップＳ５０１で取得した定型メッセージデータが表す定型メッセージを構成する音片１個につき

１個ずつの音片データを選ぶことにより得られる各組み合わせのうち、組み合わせに属する各音片データの評価値H_{XY}の総和が最大となるものを、定型メッセージを読み上げる音声を合成するための最適な音片データの組み合わせとして選択する(ステップＳ５０７)。ただし、総和を求めるために用いられる評価値H_{XY}としては、組み合わせ内での音片の接続関係を正しく反映したものが選ばれるものとする。

一方、このパーソナルコンピュータは、欠落部分識別データを生成した場合、欠落部分識別データが示す音片の読みを表す表音文字列を定型メッセージデータより抽出し、この表音文字列につき、音素毎に、配信文字列データが表す表音文字列と同様に扱って上述のステップＳ２０２～Ｓ２０３の処理を行うことにより、この表音文字列内の各表音文字が示す音声の波形を表す波形データを復元する（ステップＳ５０８）。

そして、このパーソナルコンピュータは、復元した波形データと、ステップＳ５０７で選択した組み合わせに属する音片データとを、定型メッセージデータが示す定型メッセージ内での各音片の並びに従った順序で互いに結合し、合成音声を表すデータとして出力する（ステップＳ５０９）。

なお、パーソナルコンピュータに本体ユニットＭや音片登録ユニットＲの機能を行わせるプログラムは、例えば、通信回線の掲示板（ＢＢＳ）にアップロードし、これを通信回線を介して配信してもよく、また、これらのプログラムを表す信号により搬送波を変調し、得られた変調波を伝送し、この変調波を受信した装置が変調波を復調してこれらのプログラムを復元するようにしてもよい。

そして、これらのプログラムを起動し、ＯＳの制御下に、他のアプリケーションプログラムと同様に実行することにより、上述の処理を実行することができる。

なお、ＯＳが処理の一部を分担する場合、あるいは、ＯＳが本願発明の１つの構成要素の一部を構成するような場合には、記録媒体には、その部分を除いたプログラムを格納してもよい。この場合も、この発明では、その記録媒体には、コンピュータが実行する各機能又はステップを実行するためのプログラムが格納されているものとする。

産業上利用可能性

本発明によれば、簡単な構成で高速に自然な合成音声を得るための音声選択装置、音声選択方法及びプログラムが実現される。

請求の範囲

１．音声の波形を表す音声データを複数記憶する記憶手段と、

文章を表す文章情報を入力し、各前記音声データのうちから、前記文章を構成する音片と読みが共通する音片の波形を表している音声データを索出する検索手段と、

索出された音声データのうちから、前記文章を構成するそれぞれの音片に相当する音声データを１個ずつ、互いに隣接する音片同士の境界でのピッチの差を前記文章全体で累計した値が最小となるように選択する選択手段と、

より構成されることを特徴とする音声データ選択装置。

２．選択された音声データを互いに結合することにより、合成音声を表すデータを生成する音声合成手段を更に備える、

ことを特徴とする請求項１に記載の音声データ選択装置。

３．音声の波形を表す音声データを複数記憶し、

文章を表す文章情報を入力し、各前記音声データのうちから、前記文章を構成する音片と読みが共通する音片の波形を表している音声データを索出し、

索出された音声データのうちから、前記文章を構成するそれぞれの音片に相当する音声データを１個ずつ、互いに隣接する音片同士の境界でのピッチの差を前記文章全体で累計した値が最小となるように選択する、

ことを特徴とする音声データ選択方法。

４．コンピュータを、

音声の波形を表す音声データを複数記憶する記憶手段と、

文章を表す文章情報を入力し、各前記音声データのうちから、前記文章を構成する音片と読みが共通する音片の波形を表している音声データを索出する検索手段と、

索出された音声データのうちから、前記文章を構成するそれぞれの音片に相当する音声データを１個ずつ、互いに隣接する音片同士の境界でのピッチの差を前記文章全体で累計した値が最小となるように選択する選択手段と、

して機能させるためのプログラム。

５．音声の波形を表す音声データを複数記憶する記憶手段と、

文章を表す文章情報を入力し、当該文章を構成する音片について韻律予測を行うことにより、当該音片のピッチの時間変化を予測する予測手段と、

各前記音声データのうちから、前記文章を構成する音片と読みが共通する音片の波形を表していて、且つ、ピッチの時間変化が前記予測手段による予測の結果と最も高い相関を示す音声データを選択する選択手段と、

より構成されることを特徴とする音声選択装置。

６．前記選択手段は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化との間での１次回帰を行う回帰計算の結果に基づいて、当該音声データのピッチの時間変化と前記予測手段による予測の結果との相関の強さを特定するものである、

ことを特徴とする請求項５に記載の音声選択装置。

７．前記選択手段は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変

化との間の相関係数に基づいて、当該音声データのピッチの時間変化と前記予測手段による予測の結果との相関の強さを特定するものである、

ことを特徴とする請求項５に記載の音声選択装置。

８．音声の波形を表す音声データを複数記憶する記憶手段と、

文章を表す文章情報を入力し、当該文章内の音片について韻律予測を行うことにより、当該音片の時間長、及び、当該音片のピッチの時間変化を予測する予測手段と、

前記文章内の音片と読みが共通する音片の波形を表す各々の音声データについての評価値を特定し、評価値が最も高い評価を表している音声データを選択する選択手段と、より構成されており、

前記評価値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果との相関を表す数値の関数、及び、当該音声データが表す音片の時間長と、当該音片と読みが共通する前記文章内の音片の時間長の予測結果との差の関数より得られるものである、

ことを特徴とする音声選択装置。

９．前記相関を表す数値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化との間での１次回帰により得られる１次関数の勾配からなる、

ことを特徴とする請求項８に記載の音声選択装置。

１０．前記相関を表す数値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化との間での１次回帰により得られる１次関数の切片からな

る、

ことを特徴とする請求項８に記載の音声選択装置。

１１．前記相関を表す数値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果との間の相関係数からなる、

ことを特徴とする請求項８に記載の音声選択装置。

１２．前記相関を表す数値は、音声データが表す音片のピッチの時間変化を表すデータを種々のビット数循環シフトしたものが表す関数と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果を表す関数との相関係数の最大値からなる、

ことを特徴とする請求項８に記載の音声選択装置。

１３．前記記憶手段は、音声データの読みを表す表音データを、当該音声データに対応付けて記憶しており、

前記選択手段は、前記文章内の音片の読みに合致する読みを表す表音データが対応付けられている音声データを、当該音片と読みが共通する音片の波形を表す音声データとして扱う、

ことを特徴とする請求項５乃至１２のいずれか１項に記載の音声選択装置。

１４．選択された音声データを互いに結合することにより、合成音声を表すデータを生成する音声合成手段を更に備える、

ことを特徴とする請求項５乃至１３のいずれか１項に記載の音声選択装置。

１５．前記文章内の音片のうち、前記選択手段が音声データを選択できなかった音片について、前記記憶手段が記憶する音声データを用いることなく、当該音片の波形を表す音声データを合成する欠落

部分合成手段を備え、

前記音声合成手段は、前記選択手段が選択した音声データ及び前記欠落部分合成手段が合成した音声データを互いに結合することにより、合成音声を表すデータを生成する、

ことを特徴とする請求項１４に記載の音声選択装置。

１６．音声の波形を表す音声データを複数記憶し、

文章を表す文章情報を入力し、当該文章を構成する音片について韻律予測を行うことにより、当該音片のピッチの時間変化を予測し、

各前記音声データのうちから、前記文章を構成する音片と読みが共通する音片の波形を表していて、且つ、ピッチの時間変化が前記予測手段による予測の結果と最も高い相関を示す音声データを選択する、

ことを特徴とする音声選択方法。

１７．音声の波形を表す音声データを複数記憶し、

文章を表す文章情報を入力し、当該文章内の音片について韻律予測を行うことにより、当該音片の時間長、及び、当該音片のピッチの時間変化を予測し、

前記文章内の音片と読みが共通する音片の波形を表す各々の音声データについての評価値を特定し、評価値が最も高い評価を表している音声データを選択する、ことを特徴とする音声選択方法であって、

前記評価値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果との相関を表す数値の関数、及び、当該音声データが表す音片の時間長と、当該音片と読みが共通する前記文章内の音片の時間

長の予測結果との差の関数より得られるものである、

ことを特徴とする音声選択方法。

１８．コンピュータを、

音声の波形を表す音声データを複数記憶する記憶手段と、

文章を表す文章情報を入力し、当該文章を構成する音片について韻律予測を行うことにより、当該音片のピッチの時間変化を予測する予測手段と、

各前記音声データのうちから、前記文章を構成する音片と読みが共通する音片の波形を表していて、且つ、ピッチの時間変化が前記予測手段による予測の結果と最も高い相関を示す音声データを選択する選択手段と、

して機能させるためのプログラム。

１９．コンピュータを、

音声の波形を表す音声データを複数記憶する記憶手段と、

文章を表す文章情報を入力し、当該文章内の音片について韻律予測を行うことにより、当該音片の時間長、及び、当該音片のピッチの時間変化を予測する予測手段と、

前記文章内の音片と読みが共通する音片の波形を表す各々の音声データについての評価値を特定し、評価値が最も高い評価を表している音声データを選択する選択手段として機能させるためのプログラムであって、

前記評価値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果との相関を表す数値の関数、及び、当該音声データが表す音片の時間長と、当該音片と読みが共通する前記文章内の音片の時間

長の予測結果との差の関数より得られるものである、

ことを特徴とするプログラム。

２０．音声の波形を表す音声データを複数記憶する記憶手段と、

文章を表す文章情報を入力する文章情報入力手段と、

前記文章情報が表す文章内の音片と読みが共通する部分を有する音声データを索出する検索部と、前記索出されたそれぞれの音声データを文章情報が表す文章に従って接続した際に互いに隣接する音声データ同士の関係に基づいた所定の評価基準に従って評価値を求め、出力する音声データの組み合わせを当該評価値に基づいて選択する選択手段と、を備える、

ことを特徴とする音声データ選択装置。

２１．前記評価基準は、互いに隣接する音声データ同士の関係を示す評価値を定める基準であって、前記評価値は、前記音声データが表す音声の特徴を示すパラメータ、前記音声データが表す音声を互いに結合して得られる音声の特徴を示すパラメータ、及び、発話時間長に関する特徴を示すパラメータのうち、少なくともいずれかを含む評価式に基づいて得られるものである、

ことを特徴とする請求項２０に記載の音声データ選択装置。

２２．前記評価基準は、互いに隣接する音声データ同士の関係を示す評価値を定める基準であって、前記評価値は、前記音声データが表す音声を互いに結合して得られる音声の特徴を示すパラメータを含み、また、前記音声データが表す音声の特徴を示すパラメータと発話時間長に関する特徴を示すパラメータのうち、少なくともいずれかを含む評価式に基づいて得られるものである、

ことを特徴とする請求項２０に記載の音声データ選択装置。

２３．前記音声データが表す音声を互いに結合して得られる音声の特徴を示すパラメータは、前記文章情報が表す文章内の音片と読みが共通する部分を有する音声の波形を表す音声データのうちから、前記文章を構成するそれぞれの音片に相当する音声データを１個ずつ選択した場合における、互いに隣接する音声データ同士の境界でのピッチの差に基づいて得られるものである、

ことを特徴とする請求項２１又は２２に記載の音声データ選択装置。

２４．前記評価基準は、更に音声データが表す音声との韻律予測結果との相関ないし差異を示す評価値を定める基準を含み、前記評価値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果との相関を表す数値の関数、及び／又は、当該音声データが表す音片の時間長と、当該音片と読みが共通する前記文章内の音片の時間長の予測結果との差の関数に基づいて得られるものである、

ことを特徴とする請求項２０乃至２３のいずれか１項に記載の音声データ選択装置。

２５．前記相関を表す数値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化との間での１次回帰により得られる１次関数の勾配及び／又は切片からなる、

ことを特徴とする請求項２４に記載の音声データ選択装置。

２６．前記相関を表す数値は、音声データが表す音片のピッチの時間変化と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果との間の相関係数からなる、

ことを特徴とする請求項２４又は２５に記載の音声データ選択装置。

２７．前記相関を表す数値は、音声データが表す音片のピッチの時間変化を表すデータを種々のビット数循環シフトしたものが表す関数と、当該音片と読みが共通する前記文章内の音片のピッチの時間変化の予測結果を表す関数との相関係数の最大値からなる、

ことを特徴とする請求項２４又は２５に記載の音声データ選択装置。

２８．前記記憶手段は、音声データの読みを表す表音データを、当該音声データに対応付けて記憶しており、

前記選択手段は、前記文章内の音片の読みに合致する読みを表す表音データが対応付けられている音声データを、当該音片と読みが共通する音片の波形を表す音声データとして扱う、

ことを特徴とする請求項２０乃至２７のいずれか１項に記載の音声データ選択装置。

２９．選択された音声データを互いに結合することにより、合成音声を表すデータを生成する音声合成手段を更に備える、

ことを特徴とする請求項２０乃至２８のいずれか１項に記載の音声データ選択装置。

３０．前記文章内の音片のうち、前記選択手段が音声データを選択できなかった音片について、前記記憶手段が記憶する音声データを用いることなく、当該音片の波形を表す音声データを合成する欠落部分合成手段を備え、

前記音声合成手段は、前記選択手段が選択した音声データ及び前記欠落部分合成手段が合成した音声データを互いに結合することに

より、合成音声を表すデータを生成する、

ことを特徴とする請求項２９に記載の音声データ選択装置。

３１．音声の波形を表す音声データを複数記憶し、

文章を表す文章情報を入力し、

前記文章情報が表す文章内の音片と読みが共通する部分を有する音声データを索出し、

前記索出されたそれぞれの音声データを文章情報が表す文章に従って接続した際に互いに隣接する音声データ同士の関係に基づいた所定の評価基準に従って評価値を求め、出力する音声データの組み合わせを当該評価値に基づいて選択する、

ことを特徴とする音声データ選択方法。

３２．コンピュータを、

音声の波形を表す音声データを複数記憶する記憶手段と、

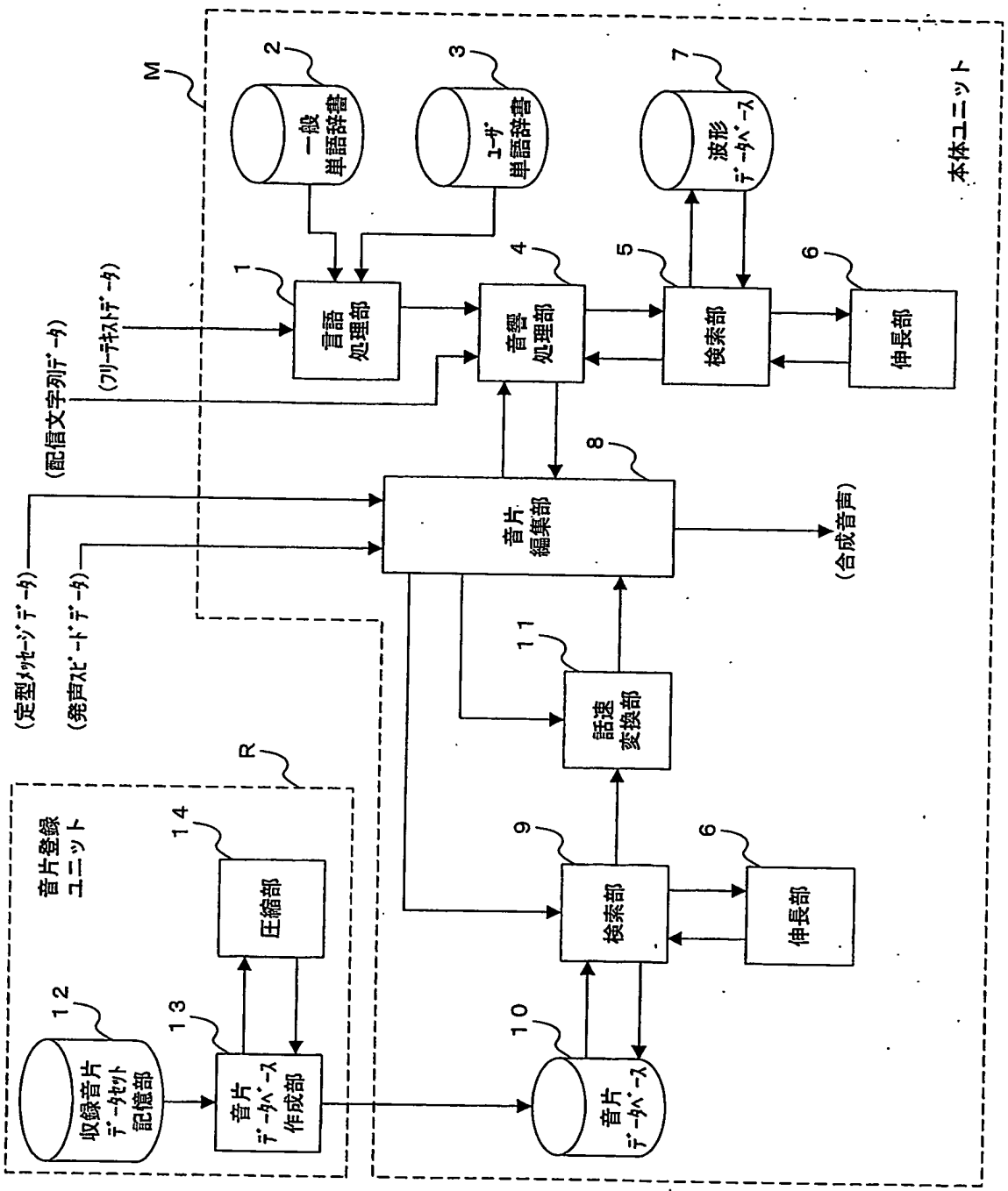
文章を表す文章情報を入力する文章情報入力手段と、

前記文章情報が表す文章内の音片と読みが共通する部分を有する音声データを索出する検索部と、

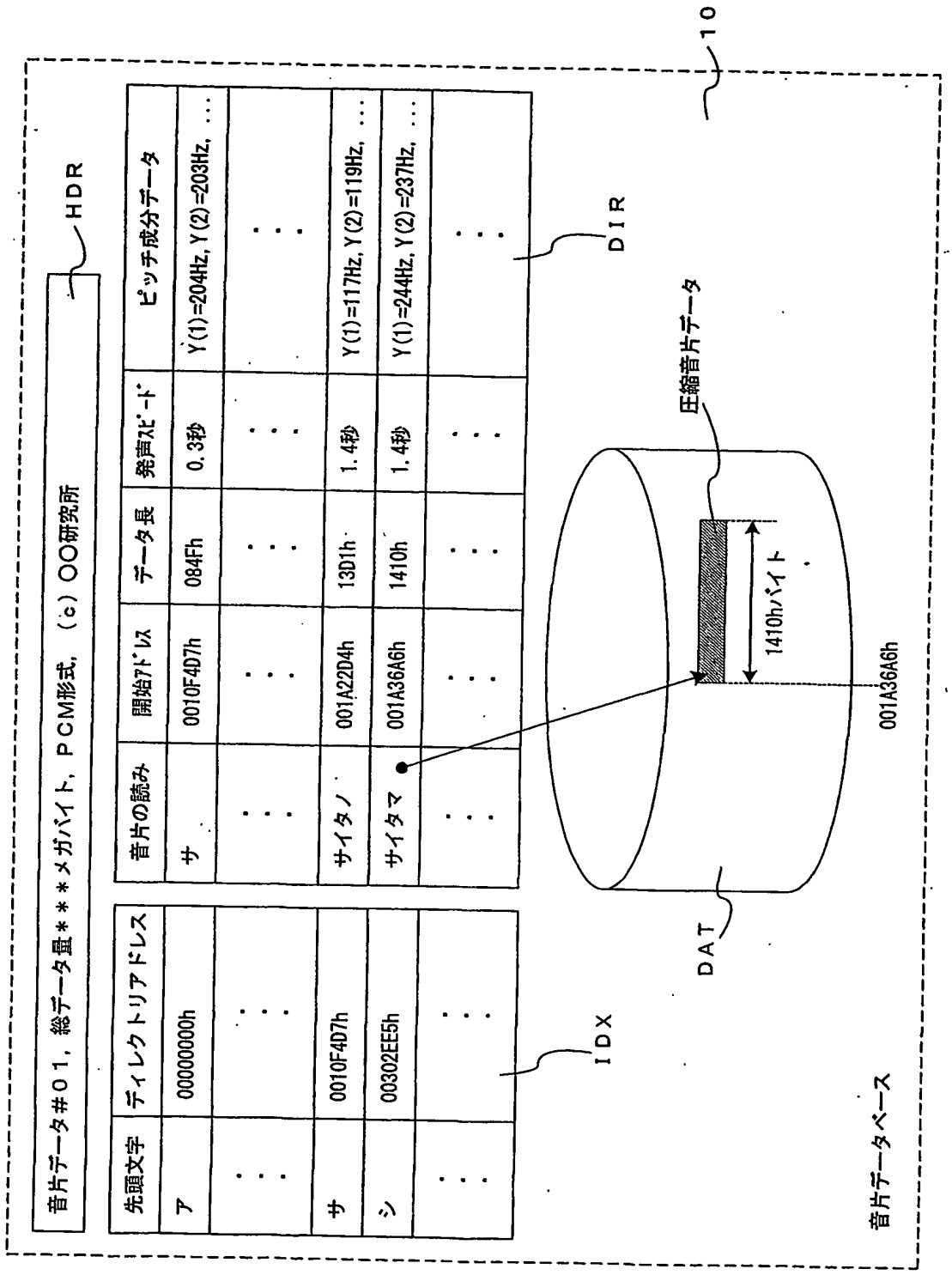
前記索出されたそれぞれの音声データを文章情報が表す文章に従って接続した際に互いに隣接する音声データ同士の関係に基づいた所定の評価基準に従って評価値を求め、出力する音声データの組み合わせを当該評価値に基づいて選択する選択手段と、

して機能させるためのプログラム。

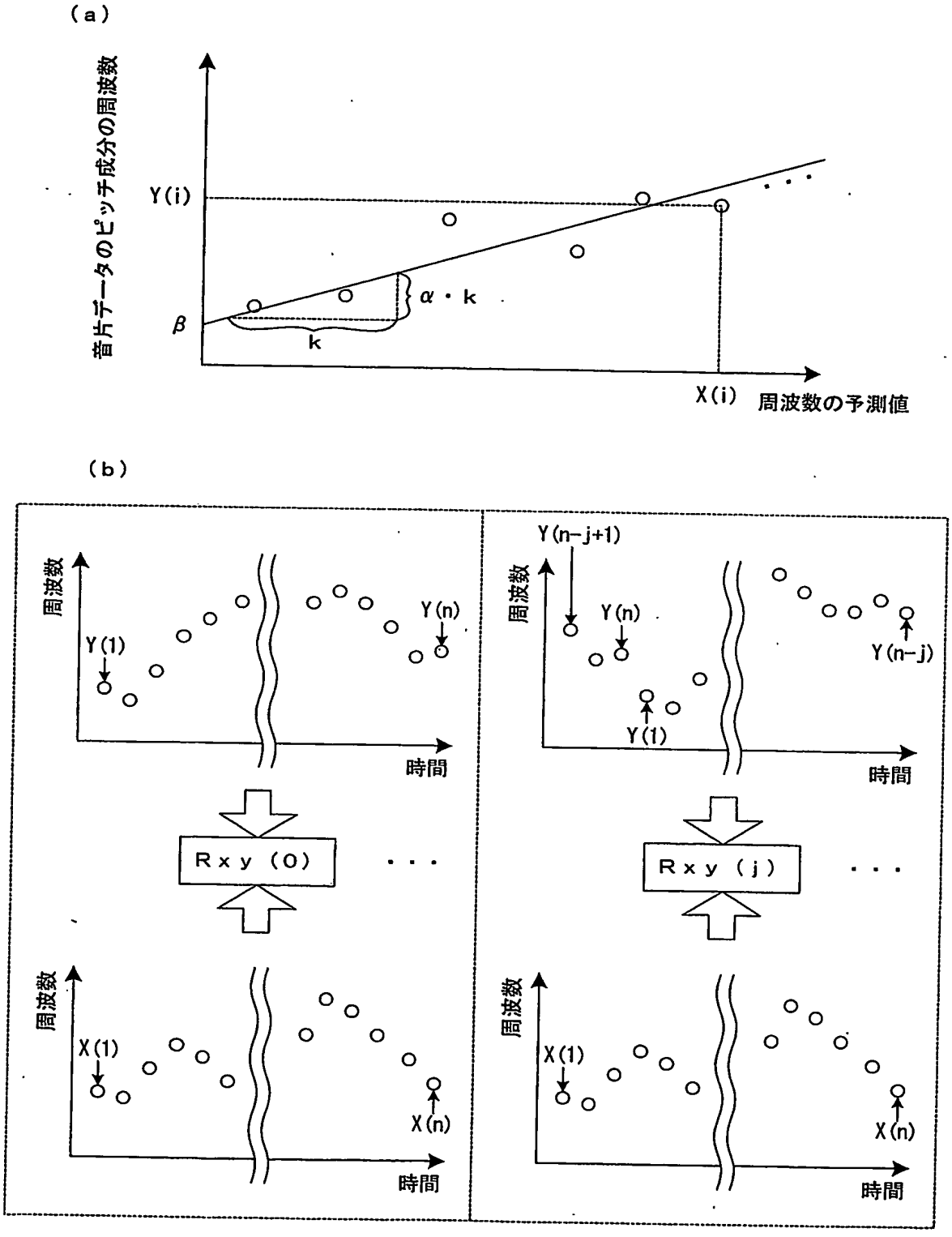
第1図


本体ユニット
M
2
一般
単語辞書
3
ユーザ
単語辞書
7
波形
データベース
1
言語
処理部
4
音響
処理部
5
検索部
6
伸長部
(配信文字列データ)
(フリーテキストデータ)
8
音片
編集部
(定型メッセージデータ)
(発声スピードデータ)
(合成音声)
11
話速
変換部
9
検索部
6
伸長部
10
音片
データベース
R
音片登録
ユニット
14
圧縮部
12
収録音片
データセット
記憶部
13
音片
データベース
作成部

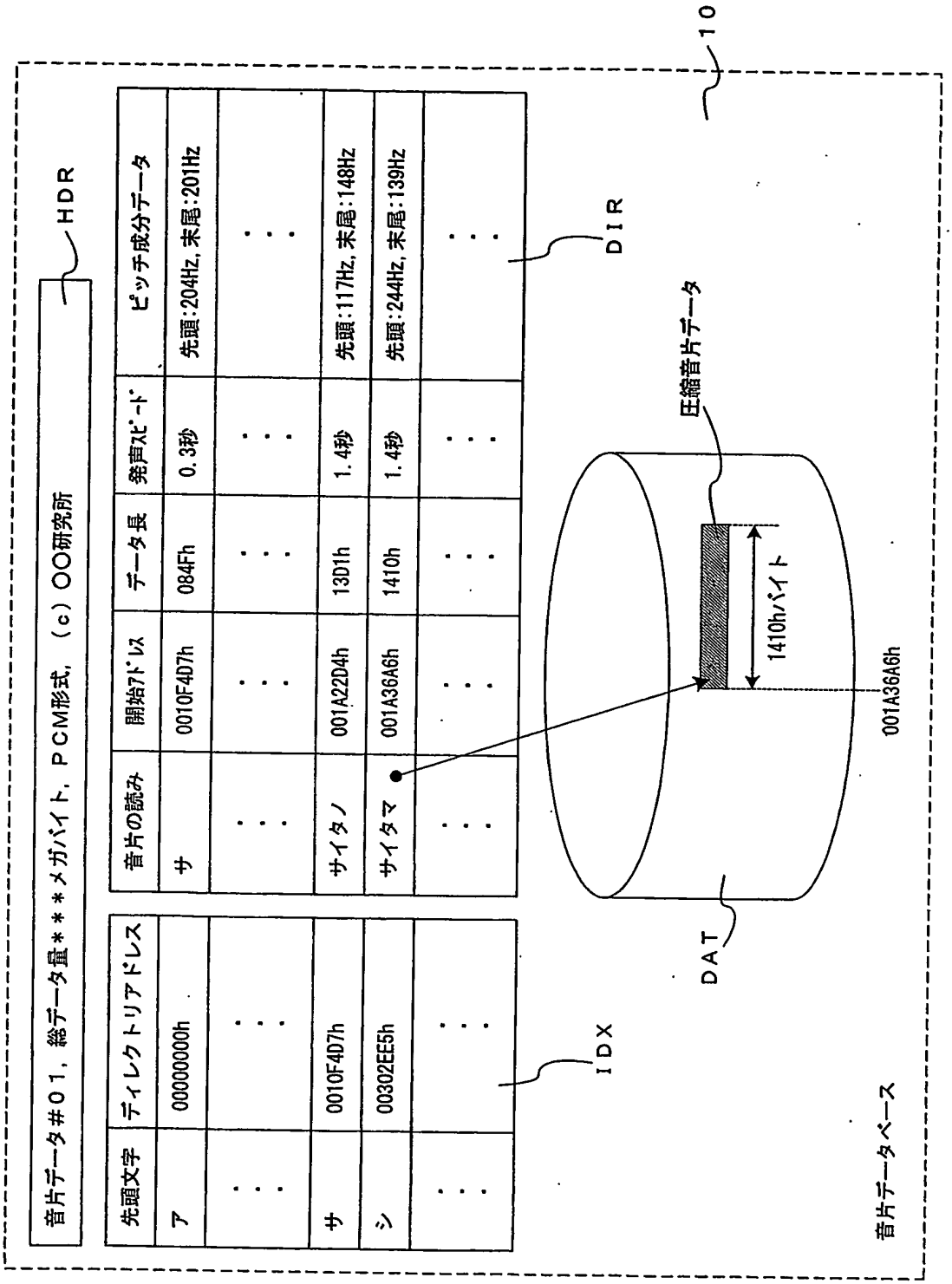
第2図


音片データ#01，総データ量＊＊＊メガバイト，PCM形式，(c) OO研究所
HDR
先頭文字
ディレクトリアドレス
ア
00000000h
サ
0010F4D7h
シ
00302EE5h
IDX
音片の読み
開始ｱﾄﾞﾚｽ
データ長
発声ｽﾋﾟｰﾄﾞ
ピッチ成分データ
サ
0010F4D7h
084Fh
0.3秒
Y(1)=204Hz, Y(2)=203Hz, ...
サイタノ
001A22D4h
13D1h
1.4秒
Y(1)=117Hz, Y(2)=119Hz, ...
サイタマ
001A36A6h
1410h
1.4秒
Y(1)=244Hz, Y(2)=237Hz, ...
DIR
10
圧縮音片データ
1410hバイト
001A36A6h
DAT
音片データベース

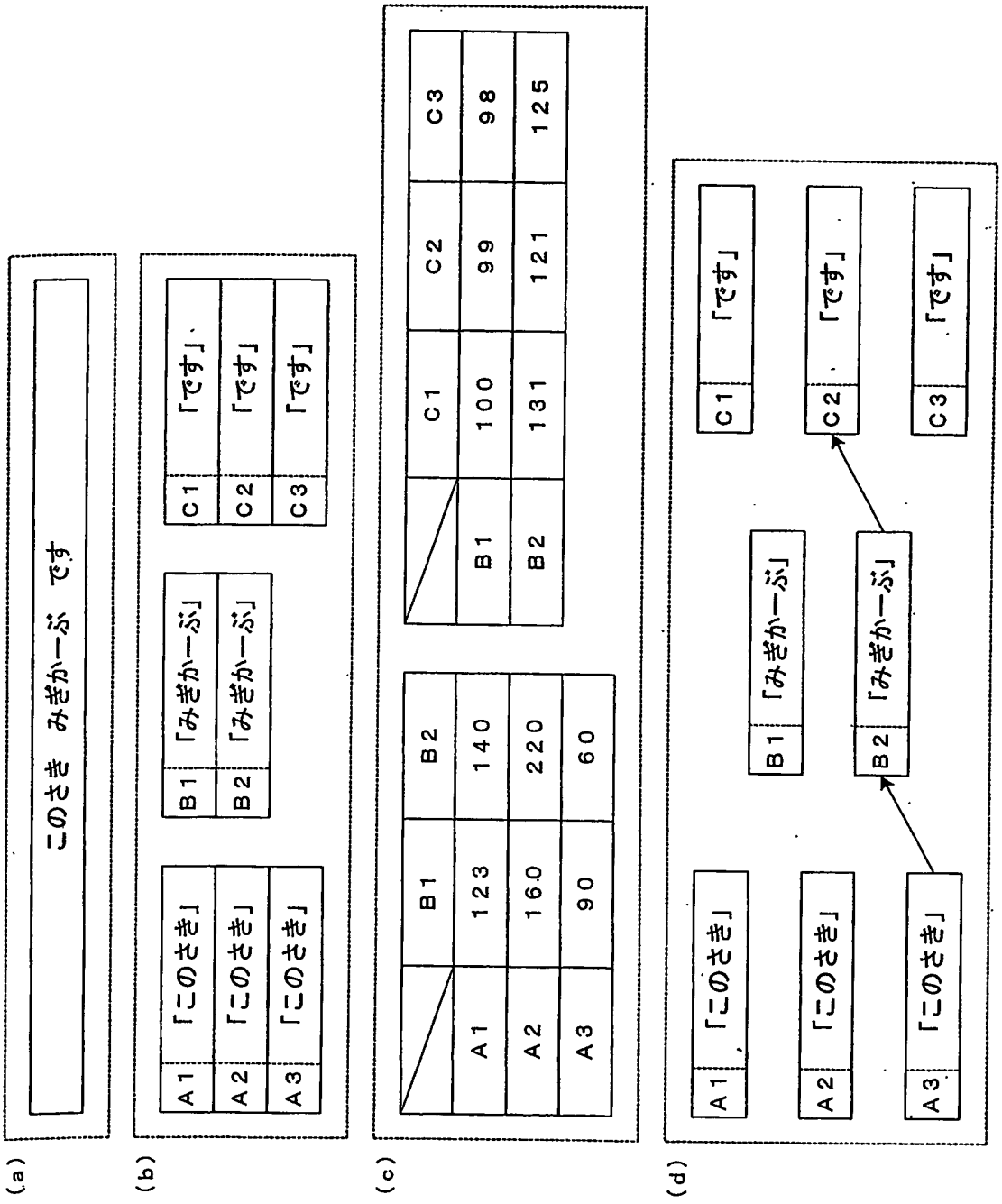
第3図


(a)
音片データのピッチ成分の周波数
Y(i)
β
α・k
k
X(i)
周波数の予測値
(b)
周波数
Y(1)
Y(n)
時間
Rxy(0)
X(1)
X(n)
Y(n-j+1)
Y(n-j)
Rxy(j)

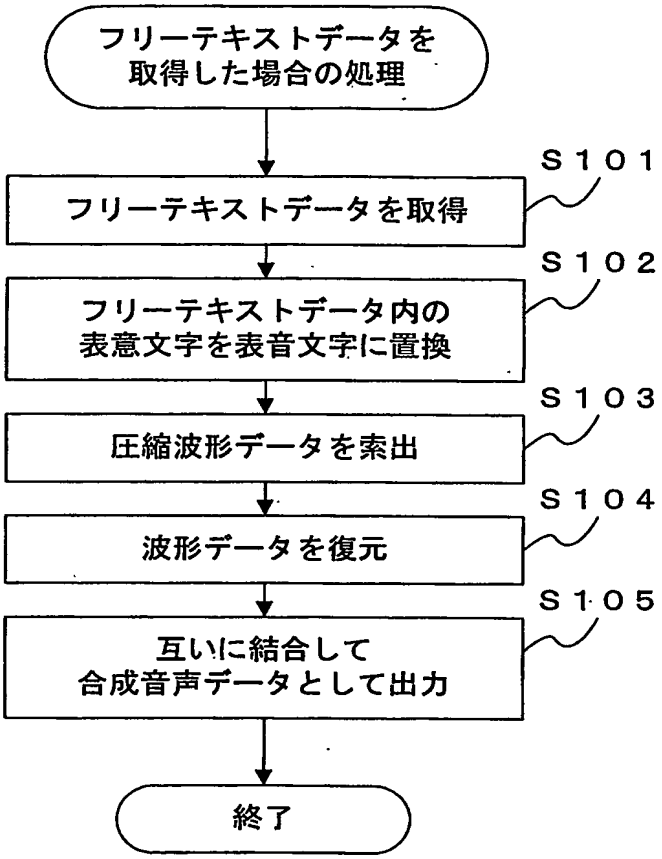
第4図


音片データ#01，総データ量＊＊＊メガバイト，PCM形式，（c）OO研究所
HDR
先頭文字
ディレクトリアドレス
ア
00000000h
サ
0010F4D7h
シ
00302EE5h
IDX
音片の読み
開始アドレス
データ長
発声スピード
ピッチ成分データ
サ
0010F4D7h
084Fh
0.3秒
先頭:204Hz, 末尾:201Hz
サイタノ
001A22D4h
13D1h
1.4秒
先頭:117Hz, 末尾:148Hz
サイタマ
001A36A6h
1410h
1.4秒
先頭:244Hz, 末尾:139Hz
DIR
10
圧縮音片データ
1410hバイト
001A36A6h
DAT
音片データベース

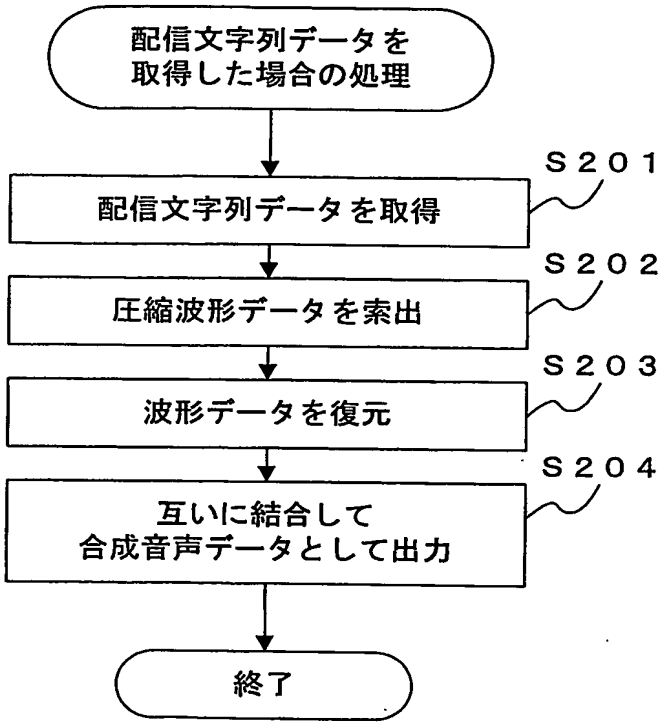
第5図


(a)
このさき みぎかーぶ です
(b)
A1 「このさき」
A2 「このさき」
A3 「このさき」
B1 「みぎかーぶ」
B2 「みぎかーぶ」
C1 「です」
C2 「です」
C3 「です」
(c)
B1 B2
A1 123 140
A2 160 220
A3 90 60
C1 C2 C3
B1 100 99 98
B2 131 121 125
(d)
A1 「このさき」
A2 「このさき」
A3 「このさき」
B1 「みぎかーぶ」
B2 「みぎかーぶ」
C1 「です」
C2 「です」
C3 「です」

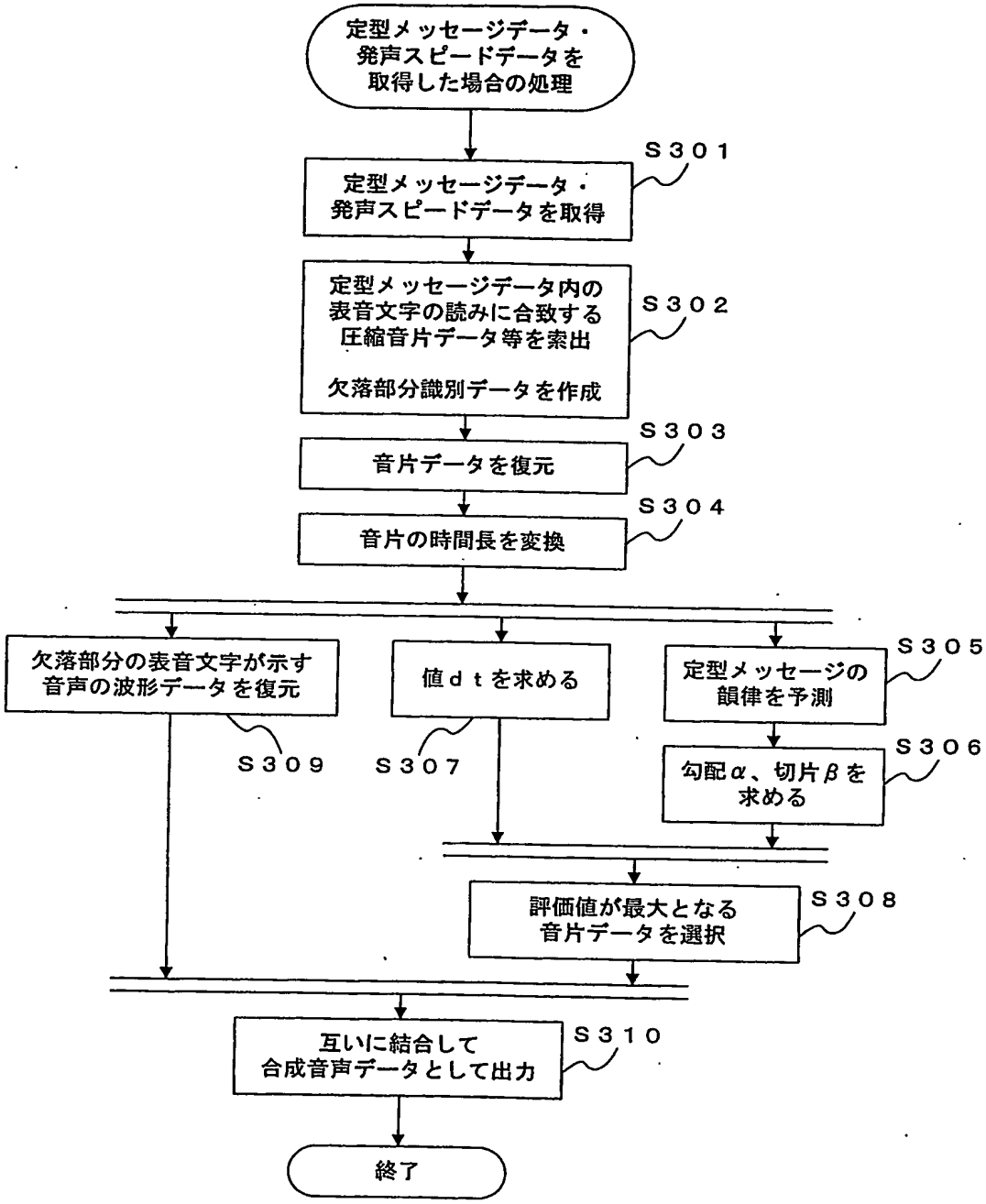
第6図


フリーテキストデータを
取得した場合の処理
S101
フリーテキストデータを取得
S102
フリーテキストデータ内の
表意文字を表音文字に置換
S103
圧縮波形データを索出
S104
波形データを復元
S105
互いに結合して
合成音声データとして出力
終了

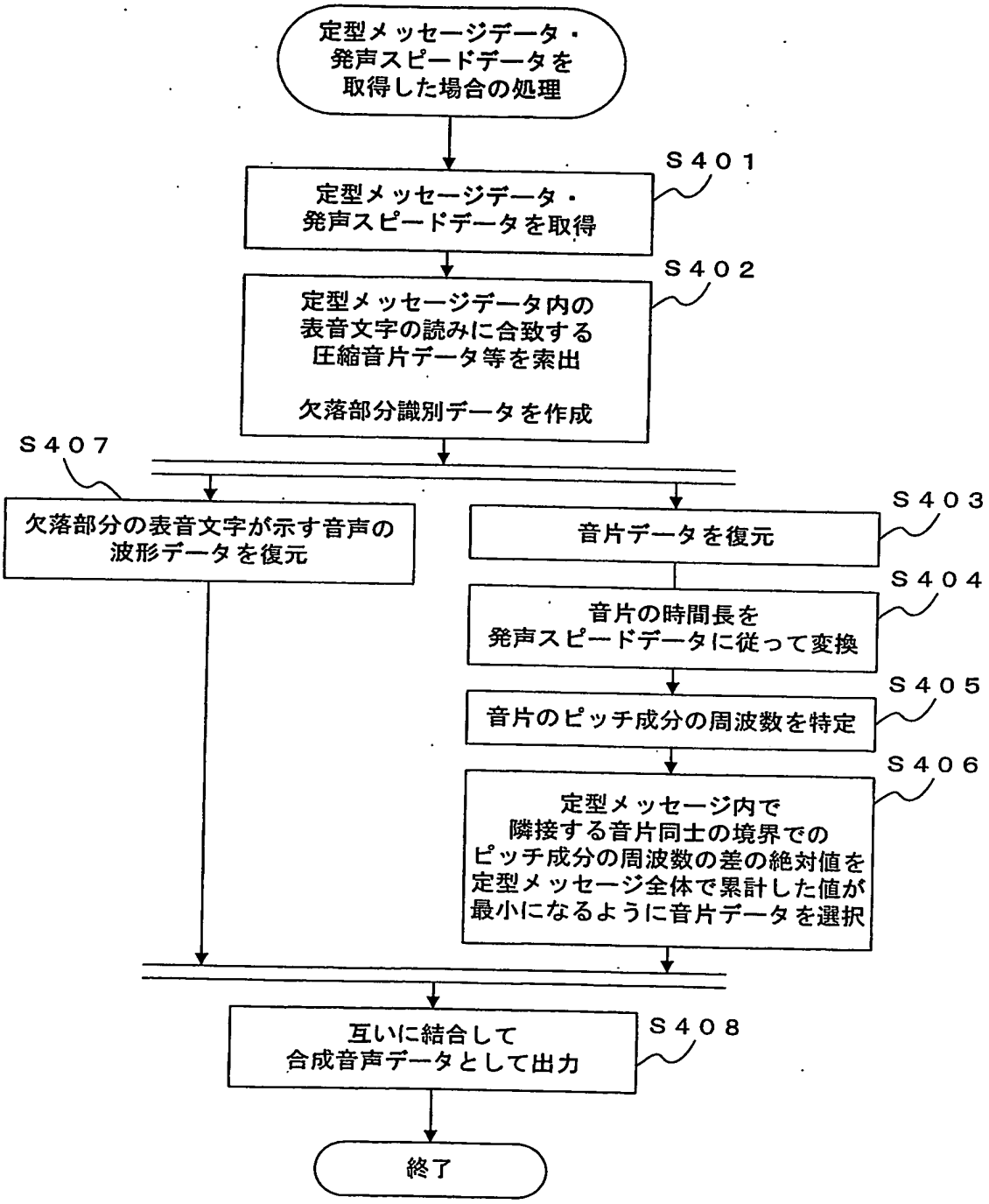
第7図


配信文字列データを
取得した場合の処理
S201
配信文字列データを取得
S202
圧縮波形データを索出
S203
波形データを復元
S204
互いに結合して
合成音声データとして出力
終了

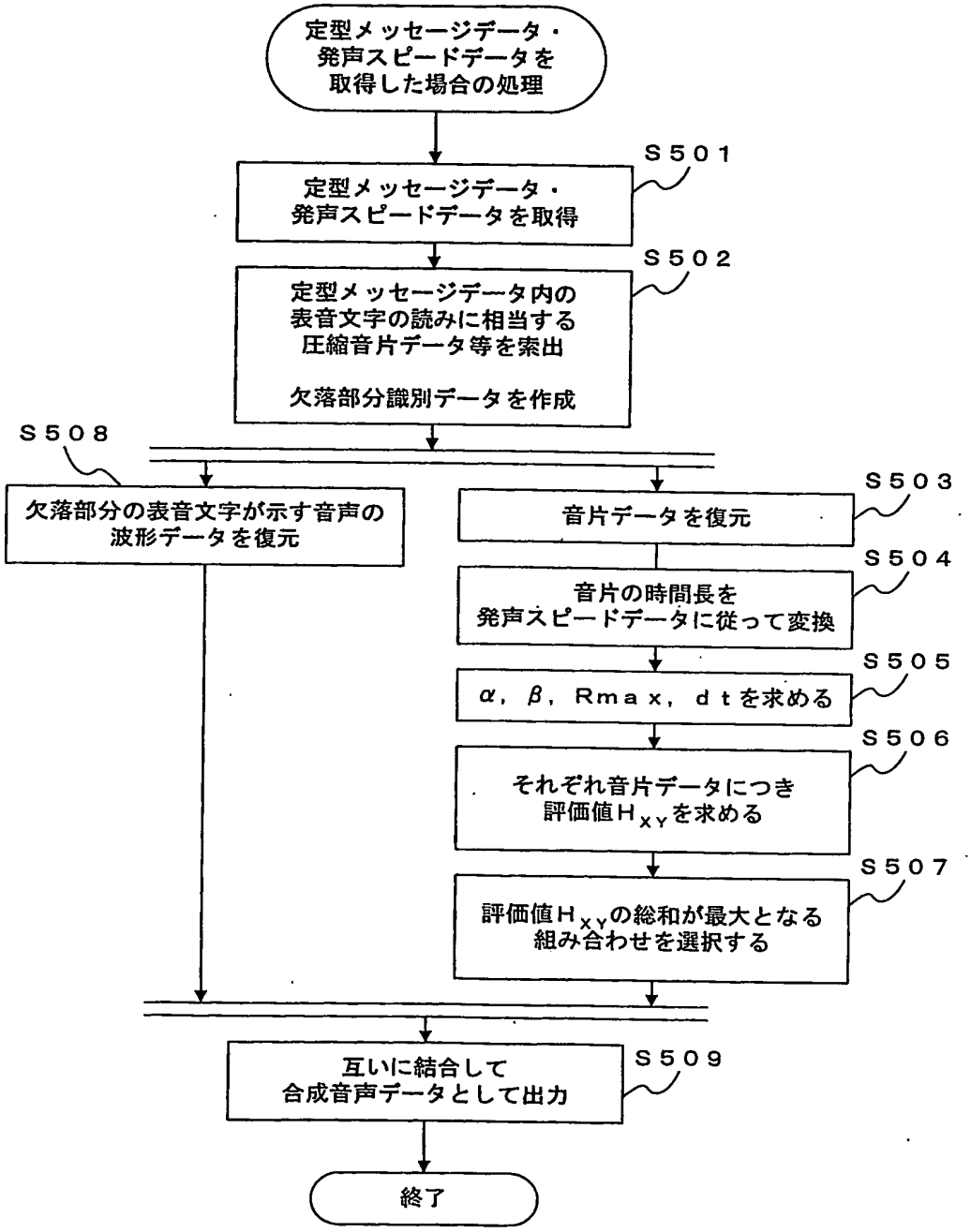
第8図


定型メッセージデータ・発声スピードデータを取得した場合の処理
S301
定型メッセージデータ・発声スピードデータを取得
S302
定型メッセージデータ内の表音文字の読みに合致する圧縮音片データ等を索出
欠落部分識別データを作成
S303
音片データを復元
S304
音片の時間長を変換
欠落部分の表音文字が示す音声の波形データを復元
S309
値ｄｔを求める
S307
S305
定型メッセージの韻律を予測
S306
勾配α、切片βを求める
S308
評価値が最大となる音片データを選択
S310
互いに結合して合成音声データとして出力
終了

第９図


定型メッセージデータ・
発声スピードデータを
取得した場合の処理
S４０１
定型メッセージデータ・
発声スピードデータを取得
S４０２
定型メッセージデータ内の
表音文字の読みに合致する
圧縮音片データ等を索出
欠落部分識別データを作成
S４０７
欠落部分の表音文字が示す音声の
波形データを復元
S４０３
音片データを復元
S４０４
音片の時間長を
発声スピードデータに従って変換
S４０５
音片のピッチ成分の周波数を特定
S４０６
定型メッセージ内で
隣接する音片同士の境界での
ピッチ成分の周波数の差の絶対値を
定型メッセージ全体で累計した値が
最小になるように音片データを選択
互いに結合して
合成音声データとして出力
S４０８
終了

第１０図


定型メッセージデータ・
発声スピードデータを
取得した場合の処理
S501
定型メッセージデータ・
発声スピードデータを取得
S502
定型メッセージデータ内の
表音文字の読みに相当する
圧縮音片データ等を索出
欠落部分識別データを作成
S508
欠落部分の表音文字が示す音声の
波形データを復元
S503
音片データを復元
S504
音片の時間長を
発声スピードデータに従って変換
S505
α，β，Rmax，dtを求める
S506
それぞれ音片データにつき
評価値H_{XY}を求める
S507
評価値H_{XY}の総和が最大となる
組み合わせを選択する
S509
互いに結合して
合成音声データとして出力
終了

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2004/008088

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl[7] G10L13/06

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
Int.Cl[7] G10L13/00-13/08

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Jitsuyo Shinan Koho 1922-1996 Jitsuyo Shinan Toroku Koho 1996-2004
Kokai Jitsuyo Shinan Koho 1971-2004 Toroku Jitsuyo Shinan Koho 1994-2004

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 2003-513311 A (Siemens AG.), 08 April, 2003 (08.04.03), Full text; Figs. 1 to 6 & WO 01/31434 A2 & EP 1224531 A2	1-32
A	JP 9-44191 A (Sanyo Electric Co., Ltd.), 14 February, 1997 (14.02.97), Full text; Figs. 1 to 9 (Family: none)	1-32
A	JP 10-97268 A (Sanyo Electric Co., Ltd.), 14 April, 1998 (14.04.98), Full text; Figs. 1 to 4 (Family: none)	1-32

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
23 August, 2004 (23.08.04)	07 September, 2004 (07.09.04)
Name and mailing address of the ISA/ Japanese Patent Office Facsimile No.	Authorized officer Telephone No.

Form PCT/ISA/210 (second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2004/008088

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 9-230893 A (NTT Data Communications Systems Corp.), 05 September, 1997 (05.09.97), Full text; Figs. 1 to 9 (Family: none)	1-32
A	JP 2001-34284 A (Toshiba Corp.), 09 February, 2001 (09.02.01), Full text; Figs. 1 to 6 (Family: none)	1-32
A	JP 1-284898 A (Nippon Telegraph And Telephone Corp.), 16 November, 1989 (16.11.89), Full text; Figs. 1 to 3 (Family: none)	1-32
A	JP 7-319497 A (NTT Data Communications Systems Corp.), 08 December, 1995 (08.12.95), Full text; Figs. 1 to 6 (Family: none)	1-32
A	JP 11-249679 A (Ricoh Co., Ltd.), 17 September, 1999 (17.09.99), Full text; Figs. 1 to 2 (Family: none)	1-32
A	JP 11-259083 A (Canon Inc.), 24 September, 1999 (24.09.99), Full text; Figs. 1 to 8 (Family: none)	1-32
A	JP 2001-13982 A (Victor Company Of Japan, Ltd.), 19 January, 2001 (19.01.01), Full text; Figs. 1 to 11 (Family: none)	1-32
A	JP 2001-92481 A (Sanyo Electric Co., Ltd.), 06 April, 2001 (06.04.01), Full text; Figs. 1 to 5 (Family: none)	1-32

国際調査報告

国際出願番号　PCT／JP2004／008088

A.　発明の属する分野の分類（国際特許分類（IPC））

Int. Cl⁷　G10L13／06

B.　調査を行った分野

調査を行った最小限資料（国際特許分類（IPC））

Int. Cl⁷　G10L13／00－13／08

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報　1922－1996年
日本国公開実用新案公報　1971－2004年
日本国実用新案登録公報　1996－2004年
日本国登録実用新案公報　1994－2004年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C.　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求の範囲の番号
X	JP 2003－513311 A（シーメンス　アクチエンゲゼルシャフト），2003.04.08，全文，図1－6 & WO 01/31434 A2 & EP 1224531 A2	1－32
A	JP 9－44191 A（三洋電機株式会社）1997.02.14，全文，図1－9（ファミリーなし）	1－32
A	JP 10－97268 A（三洋電機株式会社）1998.04.14，全文，図1－4（ファミリーなし）	1－32

☒ C欄の続きにも文献が列挙されている。

☐ パテントファミリーに関する別紙を参照。

＊　引用文献のカテゴリー
「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
23. 08. 2004	07. 9. 2004

国際調査機関の名称及びあて先	特許庁審査官（権限のある職員）		
日本国特許庁（ISA／JP） 郵便番号100－8915 東京都千代田区霞が関三丁目4番3号	山下　剛史	5C	8946
	電話番号　03－3581－1101　内線　3540		

様式PCT／ISA／210（第2ページ）（2004年1月）

C（続き）．　関連すると認められる文献

引用文献のカテゴリー＊	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求の範囲の番号
A	JP 9－230893 A（エヌ・ティ・ティ・データ通信株式会社），1997.09.05，全文，図1－9（ファミリーなし）	1－32
A	JP 2001－34284 A（株式会社東芝） 2001.02.09，全文，図1－6（ファミリーなし）	1－32
A	JP 1－284898 A（日本電信電話株式会社） 1989.11.16，全文，第1－3図（ファミリーなし）	1－32
A	JP 7－319497 A（エヌ・ティ・ティ・データ通信株式会社），1995.12.08，全文，図1－6（ファミリーなし）	1－32
A	JP 11－249679 A（株式会社リコー） 1999.09.17，全文，図1－2（ファミリーなし）	1－32
A	JP 11－259083 A（キヤノン株式会社） 1999.09.24，全文，図1－8（ファミリーなし）	1－32
A	JP 2001－13982 A（日本ビクター株式会社） 2001.01.19，全文，図1－11（ファミリーなし）	1－32
A	JP 2001－92481 A（三洋電機株式会社） 2001.04.06，全文，図1－5（ファミリーなし）	1－32